PHYLOGEOGRAPHY AND POPULATION GENETICS OF A BERINGIAN ENDEMIC: *DALLIA* (ESOCIFORMES: TELEOSTEI)

By

Matthew A. Campbell

RECOMMENDED:

Dr. Matthew Olson, Committee Member

Dr. J. Andrés López, Advisory Committee Co-Chair

Dr. Naoki Takebayashi, Advisory Committee Co-Chair

Dr. Christa Mulder, Chair, Department of Biology and Wildlife

APPROVED:

Dr. Paul Layer, Dean, College of Natural Science and Mathematics

Dr. Lawrence Duffy, Dean, Graduate School

July 29, 2011

Date

PHYLOGEOGRAPHY AND POPULATION GENETICS OF A BERINGIAN ENDEMIC: *DALLIA* (ESOCIFORMES: TELEOSTEI)

A

THESIS

Presented to the Faculty

of the University of Alaska Fairbanks

in Partial Fulfillment of the Requirements

for the Degree of

MASTER OF SCIENCE

By

Matthew A. Campbell, B.A.

August 2011

Abstract

In this thesis I examine the population genetics of an endemic Beringian freshwater fish genus, *Dallia* (blackfish). The current distribution of blackfish was heavily influenced by paleoclimatic instability during the Pleistocene. Beringian paleoclimatic changes during the Pleistocene included the fluctuating growth and decline of glaciers and an overall decrease in temperature and increased aridity in areas not adjacent to the Bering Sea. Pleistocene glacial advances resulted in the cyclical emergence of the Bering land bridge. The effects of paleoclimatic instability on blackfish distribution and abundance can be inferred through the distribution of genetic variation across the Beringian landscape. I address three basic questions: 1: Are separate populations of blackfish taxonomically distinct entities? I found that while there is clear genetic structuring and isolation, there is insufficient information to make a strong statement in this regard. 2: Did blackfish survive Pleistocene glaciations within multiple Beringian refugia? My results indicate that blackfish persisted in at least four broad geographic areas. 3: How did the Bering land bridge influence intercontinental aquatic interchange? My evidence points to close genetic relationships and potentially high exchange of blackfish across the Bering land bridge, which supports the Bering land bridge as conduit for freshwater aquatic migration.

Table of Contents

List of Figures

Page

List of Tables

Page

Acknowledgements

I would like to acknowledge my graduate committee for all their assistance in helping finish my degree program: Dr. Matthew Olson, Dr. Naoki Takebayashi, and Dr. J. Andrés López. My wife Sienna Houtte provided years of support. I was helped along by encouragement from Dr. Christine Hunter, Dr. Pat Doak, and many others during my graduate studies. Dr. Richard Boone helped me at a critical time, and deserves an immense amount of credit for my successful completion of a graduate degree in biology at the University of Alaska Fairbanks. I received support from the Department of Biology and Wildlife and Alaska Experimental Program to Stimulate Competitive Research (Alaska EPSCoR).

Chapter 1: Introduction

Cataloging biological diversity and understanding how it is distributed geographically is fundamental in both basic and applied biology. Describing and cataloging life forms has a long tradition and a wide range of applications. Understanding how biological diversity is distributed geographically can be used for determining areas best suited for conservation, how to manage utilization of biological resources, and to understand the influences of paleoclimatic variability and paleography on organisms. The thesis research described here responds to those needs as it aims to increase knowledge of a freshwater fish genus with unique biological and distributional characteristics.

The ice ages on Earth were characterized by extreme climatic variability with severe effects on biodiversity (Hewitt 2000). The Pleistocene (or Quatenary) glaciations are the most well understood ice age events due to their relatively recent occurrence (spanning the last 3 million years). The effects of the Pleistocene glaciations were least felt near the equator and more dramatic at higher latitudes (Hewitt 2000; Hewitt 2004). In North America, the Laurentide Ice-sheet and Cordilleran Ice-sheet isolated unglaciated parts of Alaska from the remainder of North America. During glacial extension, the reduction in sea level exposed the Bering land bridge between Alaska and Eastern Russia (Hopkins 1972). The region including the land bridge and several hundred kilometers to either side of it developed an Asian ecology due to the connections formed to Asia by the Bering land bridge (Pielou 1991). The term Beringia was coined to describe this region at the intersection of North America and Eurasia to highlight its unique assemblage of flora and fauna.

Beringia provides a unique opportunity to study past and ongoing biological connectivity between the two continents of Asia and North America. Many organisms occupied the Beringian refugium during times of glacial maxima, and subsequently dispersed far outside of Beringia post-glacially (Hewitt 2000). The ability of an organism to disperse broadly in northern latitudes is linked to survival through glacial periods due to a rapidly changing environment. As a result of the high dispersal ability observed among many Beringian organisms, differentiation between populations is expected to be reduced and observations to date generally support this expectation. These leaves open many questions regarding Beringian freshwater biogeography: How many refugia were present? Were refugia present in Kamchatka during times of glacial maxima? How well did the Bering land bridge permit the migration of freshwater organisms across the Bering Strait? In which direction did aquatic migration predominate across the Bering land bridge? When were organisms able to traverse the Bering land bridge?

An organism with biological characteristics suitable to help answer these questions had not been previously identified, partly as a result of high migratory ability in Beringian aquatic organisms. In this thesis, I attempt to answer these questions by examining genetic variability in blackfish (Esociformes, Esocidae: *Dallia*) under phylogeographic and population genetic approaches. Blackfish have great potential as a working model system for studies of Beringia because of the following reasons: blackfish are confined to Beringia, lack a strong swimming ability, are a primary freshwater fish, and have unique physiological properties. Blackfish demonstrate a high degree of genetic diversity not observed in other Beringian freshwater fishes, and can be

used as a model organism to inform biogeographic hypotheses in Beringia. In this chapter, I briefly describe the general biology of blackfish, and then outline the structure of this thesis.

Blackfish

Blackfish are small fish, rarely exceeding 20 cm in length (Morrow 1980), though they may grow in excess of 30 cm (Trent & Kubik 1974). Blackfish of VIII+ age class are caught very rarely, with fishes of a few years of age comprising the majority of fishes caught (Gudkov 1998; Aspinwall 1965; Blackett 1962). Overall the fish is roughly cylindrical in body shape with rounded fins. The dorsal and anal fins are far back along the body and are nearly opposite. The caudal fin of blackfish is not used as a primary locomotive source, with large pectoral fins being used for most movement.

Blackfish prefer slow-moving or still waters that are heavily vegetated, which may be glacial lakes with small gravel substrate, muddy bottomed thermokarst lakes, or slow-moving streams. On the Chukchi Peninsula, blackfish are most abundant in shallow thermokarst lakes that are silt bottomed and overgrown with plants (Gudkov 1998). In Alaska, blackfish are well characterized occurring in Interior Alaskan streams, lakes in the Bristol Bay region, and among tundra lakes and polygon ditches on the Arctic Coastal Plain (Blackett 1962; Aspinwall 1965; Ostdiek & Roland 1959).

The ecology and physiology of blackfish are poorly known. Only a handful of studies on this subject have been reported (Aspinwall 1965; Blackett 1962; Crawford 1974; Gudkov 1998; Ostdiek & Roland 1959). These studies have identified extensive

phenotypic variability across populations. In particular, the growth rate and size at maturity of blackfish varies greatly among habitats and locations. Rates of growth of blackfish in interior Alaska are much higher than those in Southwest Alaska in the Bristol Bay area. On average blackfish from an Interior stream were 64 mm length at age I+, 108 mm at II+ and 138 mm at age III+ (Blackett 1962). Blackett (1962) also notes that female blackfish were mature at 80 mm, at age 1 or 2. In contrast, Lake Aleknagik sampled in the Bristol Bay region produced blackfish categorized into nine (0+ to VIII+) age classes (Aspinwall 1965); which covered approximately the same range in length as the samples from Blackett (1962). Blackfish from Lake Aleknagik reach maturity at lengths between 49 and 50 mm and age 3 (Aspinwall 1965). These observations open three lines of inquiry regarding blackfish biology: 1) Why is there variability in growth rate among or between populations of blackfish? 2) What is the biological significance of differentiated phenotypes in different populations of blackfish? and; 3) Did early studies of age and growth in blackfish incorrectly suggest different growth rates (Aspinwall 1965)?

Seven different types of potential habitat have been sampled for blackfish in the Chukotka Autonomous Okrug in extreme eastern Russia (Gudkov 1998). Small thermokarst water bodies inhabited solely by blackfish to large glacial lakes that also contained Taranets char (*Salvelinus taranetzi*) and least cisco (*Coregonus sardinella*) were documented to contain blackfish. Blackfish growth in Chukotka was greatest in tundra-based thermokarst lakes, and was slower in thermokarst lakes that were also inhabited by Taranets char and least cisco (Gudkov 1998). The slowest growth rates

observed for blackfish in Chukotka were in lakes of glacial origin. The growth exhibited by blackfish in Chukotka approximates the observed growth rates of blackfish in similar conditions in Alaska. The rate of growth of blackfish typically is faster when the fish are younger, and then slows later in life (Morrow 1980).

Details of the migratory behavior of blackfish strongly support the speculation that long-distance migrations do not occur in this species (Aspinwall 1965; Blackett 1962). Observed migration seems to be either inshore or upstream in the spring coinciding with spawning and increased temperature. Spawning of blackfish occurred over a protracted period from May into July in one Alaska interior stream in contrast to a shorter spawning period restricted to late July in a Bristol Bay region lake (Aspinwall 1965; Blackett 1962). No direct observations of spawning have been recorded (Morrow 1980). The few details of blackfish diet indicate a diet primarily of small invertebrates (Gudkov 1998; Ostdiek & Roland 1959).

Blackfish are also unique among Beringian fish in their ability to obtain atmospheric oxygen through air breathing. Blackfish use a modified esophagus to survive in hypoxic waters (Crawford 1974). In contrast, fishes of the genus *Umbra* (mudminnows: Umbridae), close relatives of blackfish, use the gas bladder as an accessory breathing organ (Graham 1997). Mudminnows are documented employing different survival strategies than other fishes occupying ice covered waters, and use gas bubbles trapped under ice to obtain oxygen (Klinger et al. 1982; Magnuson et al. 1983; Magnuson et al. 1985). The Olympic mudminnow, (Esocidae: *Novumbra*) another close relative of blackfish, is suspected to breath air due to its tolerance of low dissolved

oxygen concentrations (Graham 1997). Air-breathing by blackfish contributes to summer and winter hypoxia tolerance, and consequently survival under conditions that are not suitable for most high-latitude freshwater fish species.

Blackfish are endemic to Beringia, which represents a unique distribution pattern among primary freshwater fishes (Lindsey & McPhail 1986). Blackfish prefer slow moving to still, heavily vegetated freshwaters, and can be found on the North Slope of Alaska, the Yukon River Basin, and the Western and Southwestern Alaskan coastal plain (Blackett 1962; Mecklenburg et al. 2002; Morrow 1980). In Russia, blackfish are found on the northern side of the Chukotka Peninsula at the Amguema River eastward along the coast (Balushkin & Chereshnev 1982; Chereshnev & Balushkin 1981). On the southern side of the Chukotka Peninsula, blackfish can be found as far West as Lake Achchen in the vicinity of Mechigmen Bay (Gudkov 1998). Blackfish can also be found on Bering Sea Islands between Alaska and Russia (Morrow 1980; Mecklenburg et al. 2002).

Scientific Classification and Relationships to Other Fishes

Blackfish are classified in the genus *Dallia.* It is commonly accepted that *Dallia* is a protacanthopterygian (=basal euteleost) fish genus belonging to the order Esociformes (pike) and the family Umbridae (mudminnows) (Nelson 2006). Euteleost fish comprise around 346 families, 2,935 genera, and 17,419 species (Nelson 2006). Euteleost fish are united into a monophyletic assemblage through a few synapomorphic morphological characters (Johnson & Patterson 1996). More specifically it is thought that *Dallia* should be placed in the superorder Protacanthopterygii, although the composition of and

interrelationships of Protacanthopterygii have been unstable since its inception (Greenwood et al. 1966; Fink 1984; Williams 1987; Sanford 1990; Rosen 1982; Fink & Weitzman 1982; Lauder & Liem 1983; Begle 1991; Begle 1992; Patterson 1994; Johnson & Patterson 1996; Ishiguro et al. 2003; López et al. 2004). The difficulties facing systematists relying on morphological evidence to infer relationships of the Protacanthopterygii include reduction, high modification, a mosaic distribution, and euteleost plesiomorphy of characters (Nelson 2006). There is mounting molecular evidence that places *Dallia* as a basal lineage in the Esocidae (pike family) instead of within the Umbridae (mudminnows) as it has been traditionally classified (López et al. 2000; López et al. 2004). Similarly, growing evidence from genetics places the pike order (Esociformes) as the sister of the Salmoniformes (López et al. 2000; Zaragüeta-Bagils et al. 2002; Ishiguro et al. 2003; López et al. 2004).

Up to three species of *Dallia* have been recognized. *Dallia pectoralis* was described in 1880 from Alaska (Bean 1880) while *D. delicatissima* was described from north-eastern Chukotka at about the same time (Nordenskiöld 1881). Subsequently, *D. delicatissima* was deemed a junior synonym of *D. pectoralis* (Jordan & Evermann 1896). *Dallia* remained monotypic until 1981 when a new Asian species, *Dallia admirabilis*, was described from the Amguema River of northern Chukotka based on morphological characters (Chereshnev & Balushkin 1981). The same authors then resurrected *D. delicatissima* for populations of *Dallia* on the northern side of the Chukotka Peninsula east of the Amguema River (Balushkin & Chereshnev 1982). The morphological differences between species of *Dallia* were attributed to survival in different glacial

refugia (Balushkin & Chereshnev 1982; Chereshnev & Balushkin 1981). Other populations of *Dallia* occurring in Russia and all Alaskan populations were considered by Balushkin and Chereshnev (1982) to be *D. pectoralis*. The validity of these three species in *Dallia* has been met with varying degrees of acceptance. Additionally, karyological data show that the Yukon River and North Slope populations of *D. pectoralis* within Alaska differ in chromosome number and variability of chromosome number among cells in an individual (Crossman & Rab 1996). The diagnostic morphological characters of Balushkin and Chereshnev (1982) applied to the two populations of *D. pectoralis* examined by Crossman and Rab (1996) did not show any significant differences.

The lack of morphological support for the observed karyological differences and the small sample sizes utilized by Balushkin and Chereshnev (1982) in describing species characteristics points out several issues with the current status of the taxonomy of *Dallia*. First, morphological variability of *Dallia* across its range is poorly described. Therefore, observed differences could also be part of among population variability within the species. Second, morphological variation in Arctic fishes can be attributed to rapid adaptation following glacial retreat. Parallel evolution in different phylogenetic lineages can produce convergence on similar ecotypes (i.e. benthic vs limnetic) in short order. Any observed morphological variability could have arisen on a very short (20,000 year) time scale and does not necessarily correspond to persistence in glacial refugia. The severity of the Illinoan glaciations also supports that *Dallia* survived only in East Beringia, which would not support separate refugia in Chukotka for the genus (Lindsey

& McPhail 1986). Third, while differences in karyotype are generally indicative of substantial differentiation for vertebrates; again the number of sampled populations is low (n=2) for these data. Therefore, it is unclear if the karyological data demonstrate variability among populations or evidence of speciation. However, the lack of morphological correspondence with the karyological data suggests that the diagnostic characters are not appropriate for taxonomic distinction within the genus.

History of Blackfish in Beringia

Historically, blackfish in Beringia were more broadly distributed, and it appears that paleoclimatic instability during the Pleistocene contributed to reduction in the distribution of blackfish. Two blackfish fossils are known, and both appear to represent distinct and extinct forms of the genus *Dallia*. Both fossils occur outside the range of extant blackfish. A fossil from the early Middle Pleistocene was discovered in Kolyma River lowlands approximately 800 km west of the current range of the species (Harington 1978; Lindsey & McPhail 1986). Blackfish were likely extirpated from this area during the heavy glaciations of Siberia during the Illinoan. A second fossil *Dallia* was recovered from Homer, Alaska on the Kenai Peninsula dating from the Late Miocene (Cavender 1969). The site of this fossil lies approximately 200 kilometers from the nearest *Dallia* population across part of the Gulf of Alaska and approximately 400 km from the nearest interior Alaska population. The Homer fossil is evidence that prior to the Pleistocene glaciations *Dallia* occupied a much greater range in East Beringia.

The extant distribution of blackfish includes two clear discontinuities (without considering the Bering Sea islands): the Brooks Range, which divides Alaskan

populations, and the Bering Sea, which divides Asian and North American populations. The current disjunct distribution of blackfish can be best understood in the context of Pleistocene glaciations. The history of drainage connections and glacial advances reduces in resolution each step farther back in time. The data for the Last Glacial Maximum (LGM), the Wisconsian, is therefore the clearest. Blackfish occur broadly in areas that remained unglaciated during the Wisconsian. Specifically, the North Slope of Alaska, Interior Alaska, and the former Bering land bridge and adjacent coastal plain in western and southwestern Alaska, and Eastern Russia were all unglaciated through that period. These unglaciated areas supported major drainage basins and a large area of flat, low elevation habitat suitable for blackfish dispersal (Hopkins 1972; Lindsey and McPhail 1986). One of these drainage basins consisted of the Chukchi Sea River, whose tributaries drained the northern parts of Beringia into the Arctic Ocean. In Alaska, the Chukchi Sea River's tributaries drained the Northern and Southern Seward Peninsula, Kotzebue Sound, and northern Saint Lawrence Island. The Chukchi Sea River also drained Northern Russia to the Amguema River. Correspondingly, the southern side of Beringia was drained by three large rivers into the North Pacific: the Anadyr Gulf River, the Kuskokwim River, and the Yukon River. The Anadyr Gulf River had western tributaries originating around the current Anadyr Gulf and reached eastward until the southern parts of Saint Lawrence Island. The Yukon and Kuskokwim Rivers followed their current courses, but extended across the Bering land bridge into the North Pacific.

Thesis research and organization

The second and third chapters of the thesis relate genetic variability across the geographic range of blackfish to past and current connections and the influences of paleogeography and paleoclimate on the distribution and abundance of this species. A fourth conclusion chapter summarizes the overall findings of the thesis.

Chapter 2: "The Mitochondrial Phylogeography of a Beringian Endemic: *Dallia*" utilizes mitochondrial DNA sequence data to answer the question: Did blackfish survive Pleistocene glaciations in multiple refugia within Beringia? Beringia is commonly treated as a single glacial refugium for aquatic organisms during the Pleistocene, and the data available to date generally cannot resolve greater detail. However, the geographic area that Beringia encompasses is quite large. Therefore, by focusing on an organism restricted to Beringia with low migration rate and unusual hypoxia and cold tolerance, the multiple refugia hypotheses may be profitably examined. To do so, I use mitochondrial genotypes to produce a median-joining network, a phylogenetic tree, a spatial analysis of molecular variance (SAMOVA), a mismatch analysis, and Bayesian skyline plots.

Chapter 3: "Population Genetics of *Dallia* in Beringia" uses additional data sources from nuclear DNA sequences to identify population genetic structuring among blackfish and to perform coalescent analyses to identify aspects of demographic history between identified populations of blackfish. I used data from two nuclear introns plus the mitochondrial data with clustering methods to identify separate groups of blackfish. The methods STRUCTURE, InStruct, and discriminant analysis of principal components

(DAPC) were used to identify population structuring. Demographic history of blackfish populations and contemporary population sizes, migration rates among populations, and divergence times among populations are estimated using the Isolation with Migration coalescent model.

Chapter 4: "Concluding Chapter", summarizes the findings of the thesis. It describes the high amount of mitochondrial genetic variability observed in blackfish and the implication of multiple separate areas in which blackfish survived glaciations and remain isolated today. The nuclear DNA data corroborates the mitochondrial data strongly, and the high degree of isolation among blackfish populations defined by geographic barriers is described. In addition, the relative timing of divergence among blackfish populations and the LGM is discussed.

References:

Aspinwall, N., 1965. *Spawning characteristics and early life history of the Alaskan blackfish,* Dallia pectoralis*, Bean.* M.S. Seattle: University of Washington.

Balushkin, A.V. & Chereshnev, I.A., 1982. Systematics of the genus *Dallia* (Umbridae, Esociformes). [In Russian.] *Trudy Zoologicheskogo Instituta Akademii Nauk SSSR*, 114, pp.36-56.

Bean, T.H., 1880. Descriptions of some genera and species of Alaskan fishes. *Proceedings of the United States National Museum*, 2, pp.353-359.

Bean, T.H., 1890. New fishes collected off the coast of Alaska and the adjacent region southward. In: Scientific results of explorations by the U.S. Fish Commission steamer Albatross. *Proceedings of the United States National Museum*, 13, pp.37-45.

Begle, D.P., 1992. Monophyly and relationships of the argentinoid fishes. *Copeia*, 1992, pp.350-366.

Begle, D.P., 1991. Relationships of the osmeroid fishes and the use of reductive characters in phylogenetic analysis. *Systematic Zoology*, 40, pp.33-53.

Blackett, R.F., 1962. Some phases in the life history of the Alaskan blackfish, *Dallia pectoralis*. *Copeia*, 1962, pp.124-130.

Cavender, T., 1969. An Oligocene mudminnow (family Umbridae) from Oregon with remarks on relationships with the Esocoidei. *Occasional Papers of the Museum of Zoology. Number 660*.

Chereshnev, I.A. & Balushkin, A.V., 1981. A new species of blackfish, *Dallia admirabilis sp. n.* (Umbridae, Esociformes), from the Amguema River basin (Arctic Chukotka). *Journal of Ichthyology*, 20, pp.25-30.

Crawford, R.H., 1974. Structure of an air-breathing organ and the swim bladder in the Alaska blackfish, *Dallia pectoralis* Bean. *Canadian Journal of Zoology*, 52, pp.1221-1225.

Crossman, E.J. & Rab, P., 1996. Chromosome-banding study of the Alaska blackfish, *Dallia pectoralis* (Euteleostei: Esocae), with implications for karyotype evolution and relationship of esocoid fishes. *Canadian Journal of Zoology*, 74, pp.147-156.

Fink, W.L., 1984. Basal euteleosts: relationships. In H. G. Moser et al., eds. *Ontogeny and systematics of fishes.* Lawrence: American Society of Ichthyology and Herpetology Special Publication, pp. 202-206.

Fink, W.L. & Weitzman, S.H., 1982. Relationships of the stomiiform fishes (Teleostei), with a description of *Diplophos*. *Bulletin of the Museum of Comparative Zoology*, 150, pp.31-92.

Graham, J.B., 1997. *Air-breathing fishes: Evolution, Diversity and Adaptation*, San Diego: Academic Press.

Greenwood, P.H. et al., 1966. Phyletic studies of teleostean fishes, with a provisional classification of living forms. *Bulletin of the American Museum of Natural History*, 131, pp.339-456.

Gudkov, P.K., 1998. Bering Sea *Dallia pectoralis* in the Chukchi Peninsula. *Journal of Ichthyology*, 38, pp.199-203.

Harington, C.R., 1978. Quaternary vertebrate faunas of Canada and Alaska and their suggested chronological sequence. *Syllogeus*, 15, pp.1-105.

Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. *Nature*, 405(6789), pp.907-913.

Hewitt, G., 2004. The structure of biodiversity - insights from molecular phylogeography. *Frontiers in Zoology*, 1, p.4.

Hopkins, D.M., 1972. The paleogeography and climatic history of Beringia during late Cenozoic time. *Inter-nord*, 12, pp.121-150.

Ishiguro, N.B., Miya, M. & Nishida, M., 2003. Basal euteleostean relationships: a mitogenomic perspective on the phylogenetic reality of the "Protacanthopterygii". *Molecular Phylogenetics and Evolution*, 27, pp.476 - 488.

Johnson, G.D. & Patterson, C., 1996. Relationshiips of lower euteloestean fishes. In M. L. J. Stiassny, ed. *Interrelationships of Fishes*. Academic Peess, pp. 251-328.

Jordan, D. & Evermann, B., 1896. The fishes of North and Middle America: a descriptive catalogue of the species of fish-like vertebrates found in the waters of North America, north of the Ishtmus of Panama. Part I. *Bulletin of the United States National Museum*, 47, pp.1-1240.

Klinger, S., Magnuson, J. & Gallepp, G., 1982. Survival mechanisms of the central mudminnow (*Umbra limi*), fathead minnow (*Pimephales promelas*) and brook stickleback (*Culaea inconstans*) for low oxygen in winter. *Environmental Biology of Fishes*, 7, pp.113-120.

Lauder, G.V. & Liem, K.F., 1983. The evolution and interrelationships of the actinopterygian fishes. *Bulletin of the Museum of Comparative Zoology*, 150, pp.95-197.

Lindsey, C.C. & McPhail, J.D., 1986. Zoogeography of fishes of the Yukon and McKenzie Basins. In C. H. Hocutt & E. O. Wiley, eds. *Zoogeography of the freshwater fishes of North America*. New York: Wiley Interscience, pp. 639-674.

López, J.A., Bentzen, P. & Pietsch, T.W., 2000. Phylogenetic Relationships of Esocoid Fishes (Teleostei) Based on Partial Cytochrome b and 16S Mitochondrial DNA Sequences. *Copeia*, 2000, pp.420-431.

López, J.A. et al., 2004. Esociform Phylogeny. *Copeia*, 2004, pp.449-464.

Magnuson, J.J. et al., 1983. Breathing Gas Mixtures Different from Air: An Adaptation for Survival Under the Ice of a Facultative Air-Breathing Fish. *Science*, 220, pp.312 -314.

Magnuson, J.J. et al., 1985. Surviving winter hypoxia: behavioral adaptations of fishes in a northern Wisconsin winterkill lake. *Environmental Biology of Fishes*, 14, pp.241-250.

Mecklenburg, C.W., Mecklenburg, T.A. & Thorsteinson, L.K., 2002. *Fishes of Alaska*, Bethesda, Maryland: American Fisheries Society.

Morrow, J.E., 1980. *The freshwater fishes of Alaska.*, Anchorage: Alaska Northwest Publishing Company.

Nelson, J.S., 2006. *Fishes of the World* 4th ed., Hoboken, New Jersey: John Wiley & Sons, Inc.

Nordenskiöld, A.E., 1881. *Vegas färd kring Asien och Europa*, Stockholm.

Ostdiek, J.L. & Roland, M.N., 1959. Studies on the Alaskan blackfish *Dallia pectoralis* I. Habitat, size and stomach analyses. *American Midland Naturalist*, 61, pp.218-229.

Patterson, C., 1994. Bony fishes. In D. R. Prothero & R. M. Schoch, eds. *Major features of vertebrate evolution. Short courses in paleontology.* Knoxville: Paleontological Society, pp. 57-84.

Pielou, E.C., 1991. *After the Ice Age: the return of life to glaciated North America*, Chicago: University of Chicago Press.

Rosen, D.E., 1982. Teleostean interrelationships, morphological function and evolutionary inference. *American Zoologist*, 22, pp.261-273.

Sanford, C.P.J., 1990. The phylogenetic relationships of salmonoid fishes. *Bulletin of the British Museum (Natural History). Zoology*, 56, pp.145-153.

Trent, T.T. & Kubik, S.W., 1974. *Studies of introduced blackfish in waters of southcentral Alaska.*, Alaska Department of Fish and Game.

Williams, R.R.G., 1987. *The phylogenetic relationships of the salmoniform fishes based on the suspensorium and its muscles.* Ph. D. Edmonton: University of Alberta.

Zaragüeta-Bagils, R. et al., 2002. Assessment of otocephalan and protacanthopterygian concepts in the light of multiple molecular phylogenies. *Comptes Rendus Biologies*, 325, pp.1191-1207.

CHAPTER 2: MITOCHONDRIAL PHYLOGEOGRAPHY OF A BERINGIAN ENDEMIC: *DALLIA*[1]

We examine mitochondrial genetic variability in *Dallia* (Esociformes: Esocidae) from samples across Beringia. We compare mitochondrial DNA sequence data using a phylogenetic tree, a median-joining network, Bayesian Skyline Plots, mismatch analysis, and SAMOVA to quantify relationships among the mitochondrial lineages of *Dallia* in the context of their spatial distribution. We uncovered a high degree of mitochondrial diversity that is strongly segregated into geographic areas corresponding to historical and contemporary barriers with minimal mixing of mitochondrial haplotypes between geographic areas. Mitochondrial diversity is highest in the common delta formed by the Yukon and Kuskokwim Rivers near the Bering Sea while other areas sampled in this study generally show low levels of mitochondrial variability. The observed mitochondrial diversity is consistent with persistence of mitochondrial lineages in multiple refugia through the last glacial maximum.

[1] Campbell, M.A. and J.A.L. Lopez. Unpublished. Mitochondrial Phylogeography of a Beringian Endemic: *Dallia*. Prepared for submission to Copeia.

Introduction

Pleistocene glaciations played a dominant role in shaping the present diversity, distribution, and genetic variability of the Holarctic biota (Hewitt 2000). Organisms that persisted in the Arctic and sub-Arctic through the Pleistocene were subject to extreme environmental changes affecting physical and ecological characteristics throughout the region. Among Arctic and sub-Arctic landscapes, Beringia is a distinct biogeographic region characterized by a unique faunal assemblage that is in large part the product of Pleistocene climatic instability and of its location at the intersection of North America and Eurasia. As described by Pielou (1991), Beringia comprises unglaciated parts of Alaska and the Yukon Territory, northeastern Siberia, and the Bering land bridge. East of Beringia, the Cordilleran and Laurentide ice-sheets covered much of North America, isolating Beringia from the remainder of North America. During stages of maximum glacial advance, the majority of Beringia remained unglaciated, providing a large refugium for terrestrial and freshwater organisms. Glacial advances also brought along more arid conditions, sea level drops, and the emergence of the Bering land bridge and other shallow coastal areas. The Bering land bridge provided a path for terrestrial and freshwater organisms to move between North America and Eurasia, which when combined with its isolation from North America, supports the assertion that during the Pleistocene, Beringia was ecologically most similar to Asia (Pielou 1991). At the end of the most recent glacial maximum, expansion of suitable habitat outside Beringia resulted in post-Pleistocene range expansion from Beringia for most organisms (Hewitt 2000).

The Beringian freshwater fauna was particularly affected by paleoclimatic fluctuations of the Pleistocence (Pielou 1991). Changing sea levels during the Pleistocene caused large and rapid fluctuations in the amount of available aquatic habitat (Lindsey & McPhail 1986). Ice sheets and glacial advances extirpated aquatic organisms by obliterating or severely altering habitats (Pielou 1991). Aquatic organisms could only escape from glacial advances by direct water connections that permitted dispersal away from advancing ice (Pielou 1991; Bernatchez & Wilson 1998). Debate persists regarding the environment that prevailed in unglaciated high-latitude areas. The prevailing view holds that during glacial advances the exposed continental shelf of the land bridge had a mesic climate, whereas the surrounding sea ice and glaciated mountains produced a more xeric environment in the remainder of Beringia (Hopkins 1972). In contrast, during interglacial periods the continental shelf was submerged and much of the remaining terrestrial parts of Beringia likely had a continental climate comparable to today (Hopkins 1972).

Of the many components of the freshwater fauna, fishes were particularly affected by glaciations (Bernatchez & Wilson 1998). The present composition of fish assemblages in Beringian drainages is dependent on the ability of fishes to survive in refugia, and the subsequent colonization routes available from refugia (Lindsey & McPhail 1986). Following glacial retreat, many Beringian freshwater fishes used proglacial lakes to colonize formerly glaciated areas (McAllister et al. 1986); examples include *Salvelinus namaycush* (Wilson & Hebert 1998) and *Coregonus clupeaformis* (Bernatchez & Dodson 1991; Bodaly et al. 1992). Fishes with tolerance for brackish and

saline waters were able to disperse rapidly along continental margins as well (McAllister et al. 1986). Because of these two methods of post-glacial dispersal, fishes that are thought to have persisted in Pleistocene refugia in Beringia typically have ranges that extend well beyond Beringia. However, the discrete and island-like nature of freshwater habitats allows effects of the Pleistocene glaciations to be observed in the contemporary distribution of genetic variation in freshwater fishes (e.g.,Wilson & Hebert 1998; Brunner et al. 2001; Lu et al. 2001; Turgeon & Bernatchez 2001; Stamford & Taylor 2004; Elmer et al. 2008; Harris & Taylor 2010). Beringian freshwater fishes were only able to persist in areas free from glaciers, resulting in a widespread reduction in size and number of populations with effects that are evident in the mitochondrial DNA (mtDNA) diversity of living populations (Bernatchez & Wilson 1998). Fish from glaciated areas have shallower intraspecific mtDNA clade depths and fewer mitochondrial haplotypes distributed across a wider area, when compared to fishes from non-glaciated areas (Bernatchez & Wilson 1998). Fishes with entire or partial freshwater residence show patterns of genetic diversity indicative of survival in putative refugia during glaciations, including *Thymallus arcticus* (Stamford & Taylor 2004), *Salvelinus* (Phillips et al. 1999; Brunner et al. 2001), *Oncorhyncus keta* (Seeb & Crane 1999), and *Lota lota* (Van Houdt et al. 2005; Elmer et al. 2008). However, there is low concordance in the phylogeographic patterns documented for fishes occupying previously glaciated areas (Bernatchez & Wilson 1998). Fish species with different ecologies and dispersal abilities would be affected differently by the environmental changes associated with glaciations,

which may account for the lack of phylogeographic concordance found in Beringian fishes (Bernatchez & Wilson 1998; Burridge et al. 2008).

The genus *Dallia* is unique among freshwater Beringian fishes for being presently restricted to Beringia (Lindsey & McPhail 1986) and for its presumed low dispersal ability. Up to three species are recognized in the genus (Nelson 2006). *Dallia pectoralis* (Alaska blackfish) occurs east and west of the Bering Strait, with the largest proportion of its range contained in the state of Alaska, USA. Within Alaska, *D. pectoralis* occurs along the Arctic coastal plain from the Colville River on the North Slope to the northern side of the Alaska Peninsula with a gap of undefined extent created by the topography of the Brooks Range. In interior Alaska, the range of *D. pectoralis* extends along the Yukon River drainage to the vicinity of Fairbanks on the Tanana River and in the interior Kuskokwim River drainage. Interestingly, this freshwater species is also found on Bering Sea islands that formerly were part of the Bering Land Bridge (Mecklenburg et al. 2002). In Russia, *D. pectoralis* is found along the Arctic Coast of the Chukotka Peninsula to the coastal areas of the Bristol Gulf in extreme eastern Russia (Balushkin & Chereshnev 1982; Gudkov 1998). *Dallia delicatissima* (Pilkhykay blackfish) is found on the northern side of the Chukotka Peninsula East of the Amguema River (Balushkin & Chereshnev 1982). *Dallia admirabilis*, which is considered by some as a dwarf form of *D. pectoralis* (Andreev 2004) is restricted to the Amguema River basin in Chukotka (Chereshnev & Balushkin 1981).

Here, we present the first examination of genetic diversity within and between populations in this genus to better understand the relationship between taxonomic

designations and genealogical relatedness and to guide hypotheses of how glacial cycling shaped their present diversity.

We predicted that *Dallia* would show high intra-specific genetic diversity and a strongly spatially patterned distribution of genetic diversity which would be atypical of other Beringian freshwater fish. This prediction is based on the biology of blackfish; low dispersal ability would contribute to population subdivision, while broad physiological tolerance would contribute to persistence of *Dallia* across a broader expanse of the Beringian landscape. In this paper, we describe the geographic distribution of mitochondrial DNA variability across the range of *Dallia* in Alaska and eastern Russia, including an introduced population found in the vicinity of Anchorage, Alaska. We discuss the phylogeography of *Dallia* in relation to paleoclimate and paleogeography and to other species of Beringia freshwater fishes. Given the unique distribution of *Dallia* and dearth of strictly freshwater restricted species in Beringia, our study offers unique insights on the biological impacts of glacial cycling on freshwater habitats.

Materials and Methods

We obtained samples of *D. pectoralis* from localities broadly representing the species range on the Alaska mainland, Saint Lawrence Island, and from two locations on the Chukotka Peninsula (Table 1; Fig. 1). Our sample includes populations from the North Slope of Alaska and fish from the type locality of *D. admirabilis* in Russia to gain insight into questions of taxonomic boundaries within the genus. Representatives from introduced *D. pectoralis* in the vicinity of Anchorage, Alaska are included in a attempt to identify the source population for these individuals. For this study, we examined tissue

samples from 188 individuals collected between October 2008 and June 2010. Samples were collected as frozen whole fish, whole fish with fin clips, or as fin clips. Wherever possible, we preserved whole fish in formalin and archived them at the University of Alaska Museum, Fairbanks, Alaska. All handling of fish in the study conformed to standards set forth by the University of Alaska Institutional Animal Care and Use Committee in protocol number 09-02. Fin clips were preserved in a DNA preservation buffer solution of dimethyl sulfoxide (DMSO), salt (NaCl) and 0.25M ethylenediaminetetraacetic acid (EDTA) (Seutin et al. 1991) or 95% ethanol and stored at -20° C. We also obtained archived tissue samples of Russian *D. pectoralis* and D. *admirabilis* from Siberia from the fish collections at the Burke Museum of Natural History and Culture, University of Washington, Seattle, Washington (*D. pectoralis*: UW 041670 and UW 041671; *D.admirabilis* UW041669). We made use of archived tissues of *Dallia pectoralis* (tissues: TO2653M, TO2654H, TO2655H, TO2656M, TO2657M, TO2658M, TO2659M, TO2660M, TO2689M, TO2690M, TO2690M, TO2691H, TO4305M, TO2692M, TO2693H, TO2696M, TO2697M, TO2699L, TO2700M) from the Royal Ontario Museum, Toronto, Ontario. We extracted genomic DNA from all samples with Qiagen DNEasy spin-column kits.

We designed oligonucleotide primers to target the cytochrome c oxidase I (COI) and control regions guided by the publicly available mitochondrial genome sequence from *D. pectoralis* (GenBank accession number AP004102). To verify the identity of sequenced PCR products, we performed BLAST (Benson et al. 2005) searches of one representative sequence generated in our study against the Genbank nucleotide database.

The forward COI primer (Dallia_pectoralis_COI_5444) sequence is '5-GCC ATC TTA CCT GTG GCA ATC AC-3', and the reverse COI primer Dallia_pectoralis_COI_5997) sequence is '5-AGT AAA AGG ACT GCT GTA ATC AGC-3'. The forward primer for control region amplification (Dallia_pectoralis_ControlRegion_16011) has the following sequence: '5-CCT TAC GAC TCG TTA CCC ACC-3' and the reverse primer's (Dallia_pectoralis_ControlRegion_16817) is '5-CAA AAC CGA TGC TCT TCT CTG-3'. Optimal Polymerase Chain Reaction (PCR) conditions for the COI and control region primer sets were 1X ProMega GoTaq Flexi reaction buffer, 0.2 mM dNTP's, 1.5 mM $MgCl_2$, 0.4 μM forward primer, 0.4 μM reverse primer, 0.025 U/μL GoTaq Flexi Taq polymerase, and 1μL template (variable DNA concentrations). The thermocycler profile had an initial denaturing step at 94° C for 2 min, followed by 35 cycles of denaturing at 94° C for 30 s, annealing at 52° C for 30 s, and extension at 72° C for 45 s. A final extension step at 72° C for 5 min ended the profile. Unpurified PCR products were sequenced by High-Throughput Sequencing Solutions at the University of Washington, Seattle, U.S.A. using ABI Big-Dye v3.1 chemistry in ABI 3730XL machines.

We examined and edited the raw sequence data with CodonCode Aligner version 3.0.3 (http://www.codoncode.com), and aligned cleaned and edited sequences by hand in Mesquite version 2.71 (Maddison & Maddison 2009). To concatenate COI and control region sequences, we used PhyUtility version 2.2 (Smith & Dunn 2008). Unique haplotypes from the concatenated mitochondrial DNA (mtDNA) were identified with DnaSP (Librado & Rozas 2009). Haplotype diversity was calculated with DnaSP version 5 as well. Average pairwise differences (k) and nucleotide diversity per site (π) for

concatenated mitochondrial datasets were calculated with DnaSP version 5 (Librado & Rozas 2009). We constructed a Bayesian tree from the concatenated dataset in MrBayes 3.2 (Huelsenbeck & Ronquist 2001; Ronquist & Huelsenbeck 2003), using the general time-reversible (GTR) model with four category gamma distributed (Γ) rate variation as determined by ModelTest (Posada & Crandall 1998). To visualize the degree of genetic variability and its spatial distribution, we created a median-joining network of haplotypes (Bandelt et al. 1999) using Network 4.516 (www.fluxus-engineering.com) from the concatenated mtDNA alignment.

To identify barriers and connections between sampled populations, we employed Spatial Analysis of Molecular Variance (SAMOVA) as implemented in the program SAMOVA 1.0 (Dupanloup et al. 2002). SAMOVA 1.0 uses a simulated annealing algorithm to place N populations into $K \geq 2$ groups. This is done by maximizing the proportion of total genetic variance due to differences between groups of populations (F_{ct}). We used $K=\{2,\ldots,6\}$, to create two through six groups of sampled populations with maximized F_{ct}. Program options used included 100 simulated annealing processes and pairwise differences to calculate fixation indices. SAMOVA 1.0 assigns N sampled populations to K groups and reports the genetic structure corresponding to K, and three fixation indices based on this structure: F_{ct}; F_{sc}, the proportion of genetic variance due to differences between populations within each group; and F_{st}, the genetic variance due to overall differences between populations not considering group structure. The significance level of the fixation indices is evaluated through 1,000 permutations of populations among groups.

Because we are interested in understanding the geographic and demographic effects of glacial cycling on historical populations of *Dallia*, we applied two methods used to infer current and past effective population size (N_e) from observed genetic variation. First, we conducted the Bayesian Skyline Plot (BSP) analyses implemented in BEAST version 1.5.3 (Drummond et al. 2005; Rambaut & Drummond 2007), an approach based on a coalescent model incorporating demographic changes in population size (Drummond et al. 2002). Due to the assumption that all alleles and haplotypes used in BEAST are from a panmictic population, we divided *Dallia* populations into Interior, Unalakleet-Nome, North Slope, and Coastal population groups based on the SAMOVA results, the Bayesian tree, and the geographic distribution of *Dallia*. BEAST requires unlinked loci for analysis; therefore we could not treat the COI and control region sequences separately. We used concatenated mtDNA sequences for BSP analyses, with the closest mutational model available in BEAST after selection by ModelTest (Posada & Crandall 1998). For each population, we pooled and re-sampled the results from three separate MCMC searches with Logcombiner version 1.5.3 (Rambaut & Drummond 2007). Each MCMC search was 100 million steps in length, with sampling for trees and parameters every 500 steps, and a burn-in of 10 million steps (Marko et al. 2010). We analyzed the pooled data with Tracer version 1.5.3 (Rambaut & Drummond 2007).

As an alternative method to infer changes in N_e, we performed an mtDNA mismatch analysis (Rogers & Harpending 1992) with DnaSP version 5 (Librado & Rozas 2009). Data from COI and the control region were concatenated for this analysis. For mismatch distributions, we used all samples in a single test as well as separate North

Slope, Interior, Unalakleet-Nome, and Coastal populations. We tested the populations individually for evidence of changing or constant population size and report the raggedness statistic (r). To find estimates of the time in years to intrapopulation coalescence (T), we used the estimate of intrapopulation coalescent time (τ) produced by this analysis along with generation time in years and mutation rate. A generation time of two to three years (Blackett 1962; Morrow 1980), and a conservative estimate of mutation rate for animal mitochondrial DNA, $\mu = 1 \times 10^{-6}$ per nucleotide site (Jazin et al. 1998) were used to create upper and lower estimates of intrapopulation coalescent time in years.

Results

We successfully sequenced the COI fragment for 169 fish, and for the control region we were able to sequence all but five samples (all from the Amguema River). Therefore, we sequenced both COI and control region fragments for 164 specimens of *Dallia* (Table 2). All sequences are archived in GenBank. The COI fragment alignment is 512 base pairs in length and the control region fragment alignment is 718 base pairs long. The concatenated mtDNA alignment consists of 1230 sites, of which 83 are variable. The concatenated mtDNA alignment includes 45 unique haplotypes distributed among populations (Table 3), and an overall haplotype diversity of 0.954. The average pairwise difference between sequences (k) is 11.188 (or 0.9%) for a nucleotide diversity (π) of 0.00914.

The Bayesian tree and the median-joining network analysis of the concatenated mitochondrial alignment revealed four distinct mitochondrial lineages: Interior,

Unalakleet-Nome, North Slope, and Coastal (Figs. 2 & 3). The Interior lineage consists of fish from the Tanana River drainage and the Kuskokwim River drainage upstream of the Kuskokwim Mountains (Sample locations: 1, 2, 4-6). Only one fish from outside this geographic area clusters within this lineage (Fig. 3, Sample location 12). The Interior contained eight mitochondrial haplotypes unique to that geographic area (Table 2; Fig. 3). *Dallia* from Unalakleet and Nome (Sample locations: 13 & 14) formed a unique lineage with six haplotypes not found elsewhere. The majority of North Slope blackfish (19/22) shared a common haplotype (Table 2; Fig. 3). In addition, the North Slope lineage (Sample locations: 16-20) contained four unique mitochondrial haplotypes found only in the North Slope population. The Coastal group contained haplotypes from the Lower Kuskokwim, Togiak, Bethel, Galena, St. Lawrence Island, and Russia (Sample locations: 7-12, 15, and 21-23). Lineages from these sample locations are referred to as Coastal lineages, however they constitute a paraphyletic assemblage in a midpoint rooted tree due to the exclusion of the North Slope mitochondrial lineage. The haplotype recovered from samples from the introduced population of *Dallia* from near Anchorage, Alaska (Sample location 3) groups with Coastal lineages, but has yet to be sampled from any naturally occurring population.

SAMOVA 1.0 analyses initially (K=2) separate Interior fish (Sample locations: 1, 2, and 4-6) from the remainder of sample locations. The next (K=3) grouping of populations separates those from Unalakleet and Nome (Sample locations: 13 and 14) from Interior fish and the remainder. When four groups of populations are specified (K=4), the North Slope (Sample locations 16-20) fish separate from the previously

defined two groups and the remaining sample locations. Fixation indices for the different clustering levels are given in Table 3.

Both the Bayesian tree and SAMOVA 1.0 results support four major phylogeographic units of *Dallia*. We consider sample locations 1, 2, and 4-6 to correspond to the Interior phylogeographic unit of *Dallia*. These fish occupy the Tanana River (in the Yukon drainage) and the Kuskokwim River drainage upstream of the Kuskokwim Mountains along the northern side of the Alaska Range within Alaska. The Unalakleet-Nome phylogeographic unit is composed of fish from sample locations 13 and 14, to the North and East of Norton Sound. Fish from sample locations 16-20 on the Arctic Coastal Plain of Alaska represent the North Slope phylogeographic unit. The remaining samples from the former Bering land bridge, adjacent areas of Chukotka Russia, Western Coastal Alaska, the lower Kuskokwim River, and the Yukon River belong to the Coastal phylogeographic unit. The introduced population in Wasilla, Alaska appears to have originated from somewhere within this phylogeographic unit.

A Bayesian Skyline Plot of the Coastal phylogeographic unit suggests a bottleneck of effective population size ending approximately 20,000 years ago, followed by a rapid increase (Fig. 4). The other populations lacked sufficient data to produce informative BSPs. Mismatch analyses also supported this view by indicating that phylogeographic units of *Dallia* have a history of unstable population size (r, Table 4). The estimate of intrapopulation coalescent time for *Dallia* as a whole occurs more than 1 million years ago (Table 4). Interior and Unalakleet-Nome phylogeographic units have very similar intrapopulation coalescent times. The estimate of intrapopulation coalescent

time is shallowest for the North Slope phylogeographic unit, on the scale of tens of thousands of years while the Coastal phylogeographic unit had the deepest intrapopulation coalescent time of the phylogeographic units, on the scale of hundreds of thousands of years.

Discussion

Dallia *phylogeography*

There are two salient results from this data set. Firstly, mitochondrial haplotype lineages cluster distinctly into major groups. Secondly, these mitochondrial groups correspond with general landscape features of Beringia, not with a major Asia and North America divide. The mitochondrial phylogeography of *Dallia* indicates connectivity across the Bering land bridge.

The patterns of genetic diversity we have uncovered in *Dallia* offer a contrast to results from other Beringian fishes. The lack of phylogeographic concordance among species of Arctic freshwater fishes is the product of the survival of fish in refugia during glacial periods, followed by expansion from these refugia post-glacially with differential survival and dispersal of fish genotypes from these refugia. *Dallia* is characterized by several traits not found in other Beringian fishes, which may be expected to influence observed patterns of genetic variation. *Dallia* are poor swimmers and are restricted to still or slow moving waters, with a preference for heavily vegetated areas. Observed migrations in *Dallia* consist of short movements upstream or inshore, followed by the opposite movement in the fall to deeper waters (Aspinwall 1965; Blackett 1962; Morrow 1980). Dispersal methods common to other more strongly swimming freshwater fish,

such as head stream capture, are unlikely to be exploited by *Dallia*. Coastal plain flooding is the more likely avenue for between-drainage dispersal for members of this genus. In contrast to most Beringian fishes, which display salinity tolerance or anadromous life histories, *Dallia*, as a primary freshwater fish, is unlikely to disperse across saltwater barriers. While swimming ability, dispersal ability, and salinity tolerance have not been measured in *Dallia*, an extensive study of the ecological zoogeography of *Novumbra hubbsi* (Olympic mudminnow; Esociformes: Esocidae), an ecologically similar relative of *Dallia*, suggests a low capacity for these means of range expansion (Meldrim 1968). *Novumbra* is monotypic and restricted to western Washington State, USA, with its range most likely determined by paleoclimatic processes with little evidence for any post-Pleistocene range expansion (Meldrim 1968). Species of esociform fishes with a mudminnow-type body and ecology share restricted, relictual distribution patterns (Nelson 2006). Furthermore, *D. pectoralis* shows tolerance to both summer and winter hypoxia. Air-breathing throughout the ice-free months allows *Dallia* to survive in waters that other Beringian fish cannot (Blackett 1962; Crawford 1974). *Dallia pectoralis* has been observed during ice free months in waters with dissolved oxygen levels as low as 2.3 parts per million at 7.8° C (Ostdiek & Roland 1959). *Dallia pectoralis* also use muskrat (*Ondatra zibethicus*) ice holes to gain access to air under hypoxic conditions (Armstrong 1994). It also likely that *Dallia* can survive these hypoxic conditions under ice cover without access to air and in shallow tundra ponds that freeze to the bottom; however, the mechanisms responsible for this adaptation are not known (Ultsch 1989).

The mtDNA sequence data reported here show a strong phylogeographic signal and high degree of geographic concordance among haplotype lineages. The majority of haplotypes and haplotype clusters were found only at a particular sample location or within a limited geographic range. Observed within species mitochondrial divergence is high for a Beringian freshwater fish (Bernatchez & Wilson 1998). In contrast, comparable mitochondrial phylogeography studies in *L. lota* and *Salvelinus* uncovered haplotypes that are distributed over geographic areas much greater than the entire range of *Dallia* (Seeb & Crane 1999; Stamford & Taylor 2004; Van Houdt et al. 2005; Alekseyev et al. 2009). Intraspecific mtDNA divergences among *Coregonus spp.* and *T. arcticus* are lower than those observed in *Dallia* over comparable spatial scales (Bernatchez & Dodson 1990; Bernatchez & Dodson 1991; Lu et al. 2001; Turgeon & Bernatchez 2001; Stamford & Taylor 2004; Harris & Taylor 2010). The high mtDNA intraspecific diversity in *Dallia* is likely a result of large effective population size and lack of migration between populations.

Low or zero migration rates across biogeographic barriers due to habitat requirements and low dispersal ability of *Dallia* is best exemplified by the distribution of mtDNA haplotypes observed along the Kuskokwim River drainage (Sample locations 4-8). Haplotypes recovered from fish upstream of the Kuskokwim Mountains (Sample locations 4-6) belong to the Interior phylogeographic unit, while those sampled from locations downstream of that feature (Sample locations 7 & 8) are from the Coastal phylogeographic unit. This pattern indicates that there is no or very low effective female migration along the Kuskokwim River. No known aspect of their biology suggests the

case would be different for males. The nearest Coastal and Interior *Dallia* populations sampled in the Kuskokwim Basin are separated by less than 300 river kilometers (km); however the most closely related sampled populations to Interior Kuskokwim fish are over 1,000 river km away, and require inferring a historical connection between the Yukon and Kuskokwim basins. A similar relationship has been observed in *O. kisutch* with the use of landscape genetics techniques (Olsen et al. 2011).

The median-joining network of *Dallia* mtDNA haplotypes shows many haplotypes that are not all closely related. This pattern contrasts the expectation of a central ancestral haplotype comprising the majority of sampled fish with a few closely related haplotypes present which would be consistent with recent expansion from a single refugium. The observed pattern is indicative of population substructure. The Bayesian Skyline Plot analyses suggest the Coastal population underwent a recent bottleneck. Given this inferred history and the limitations of BSP analyses, inferences of effective population size prior to the bottleneck are not reliable. Mismatch analyses support past changes in effective population size for populations of *Dallia,* and an intrapopulation coalescence time of 1.2 to 1.8 million years before present.

Our data provide strong indication that *Dallia* survived in multiple refugia within Beringia throughout the Wisconsin period. Starting with that inference and assuming a mitochondrial divergence rate of 0.5 – 1% per million years (Nabholz et al. 2008), all sampled *Dallia* mtDNA lines diverged within the last two million years. The mtDNA variability indicates that several region-specific mitochondrial lineages exist with deep intrapopulation coalescent times preceding the LGM (12,000 years), therefore we can

tentatively reject the scenario of a post-LGM expansion from a single refugium. Among the *Dallia* populations we examined here, only the North Slope phylogeographic unit yielded evidence of a recent population expansion. The North Slope phylogeographic unit experienced an increase in effective population size starting between 30.3 and 45.5 kyr ago, corresponding to the Wisconsin glacial period. It is possible that *Dallia* were able to access the North Slope through drainages of the Chukchi Sea River during the Wisconsin. The Chukchi Sea River drained northern Alaska, St. Lawrence Island, Siberia, the northern Chukotka Peninsula to the Amguema River in the west, the northern and southern Seward Peninsula, and Kotzebue Sound flowing north across the Bering land bridge (Lindsey & McPhail 1986). The Bering land bridge and expanded coastal plain habitat provided contiguous low gradient habitat along the western side of what is now Alaska, allowing the Brooks Range to be circumvented. The North Slope *Dallia* mitochondrial lineage is nested within the clade containing Coastal haplotypes, including those recorded from Saint Lawrence Island. The historical Chukchi Sea River drained the southern side of the Seward Peninsula, which is represented in this study (Sample location 14). Sample location 14 forms part of the Unalakleet-Nome phylogeographic unit and appears to be distinct from the Coastal phylogeographic unit. This evidence does not exclude the possibility of population connectivity of *Dallia* across the northern side of Beringia. Assuming that the Bering land bridge contained suitable habitat for dispersal and occupancy of *Dallia*, large delta regions and a wetter climate than other parts of Beringia as mentioned above, the multiple cycles of fluctuating sea level eliminated and restored a large portion of the range of *Dallia* repeatedly. Following the rise of the

Bering Sea, surviving mitochondrial lineages from the Bering land bridge would be able to survive in adjacent suitable habitat. However, these lineages would likely be a random subsample of the lineages from the Bering land bridge. As a result, genetic relationships may not necessarily reflect other processes (i.e. dispersal through the Chukchi Sea River).

The observed level of genetic diversity in the North Slope phylogeographic unit suggests that *Dallia* populations survived through the Wisconsin in that region. Estimates of intrapopulation coalescent time are much greater than what would be expected for a post-LGM colonization of the Arctic Coastal Plain. The current physical barriers posed by the Brooks Range and the Bering Sea also argue against the current landscape being conducive to *Dallia* dispersal to the Arctic Coastal Plain of Alaska. The more recent colonization of the Arctic Coastal Plain of Alaska by *Dallia* is supported by the greater severity of pre-Wisconsin glaciations and the glacial-mediated range reduction of *Dallia* in Beringia. Evidence of past occurrences of *Dallia* outside its present range include a fossil from the Kolyma River, 800 km west of the current western distribution limit, and a fossil from the Kenai Peninsula in Southcentral Alaska approximately 250 km from the nearest natural population in Alaska (Cavender 1969; Harington 1978; Lindsey & McPhail 1986). The Siberian fossil is from the Early Middle Pleistocene (781 kyr ago) and Lindsey & McPhail (1986) assert that is unlikely that *Dallia* survived the Illinoian glaciation outside of East Beringia. Likewise, the fossil of *Dallia* from the Kenai Peninsula is dated to the late Miocene approximately 5 mya (Cavender 1969). Survival of *Dallia* on the North Slope during the entire Pleistocene is unlikely due to the severity

of climatic conditions, and would result in much deeper mitochondrial divergences than what has been observed due to the long time period involved.

Taxonomic Implications

Studies of morphological diversity of *Dallia* have led to the recognition of up to three species in the genus (Nelson 2006). *Dallia pectoralis* was described in 1880 from western Alaska, and *D. delicatissima* was described in 1881 from Northern Chukotka. These two species were synonymized (Jordan and Evermann 1896), and *Dallia* remained monotypic based on morphological evidence until 1981when *D. admirabilis* was described from the Amguema River (Chereshnev & Balushkin 1981). *Dallia delicatissima* was also resurrected by the same authors and a set of diagnostic morphological characters were proposed (Balushkin & Chereshnev 1982). The mtDNA haplotypes recovered from *D. admirabilis* specimens in this study showed very little divergence from other populations of *Dallia* and did not form a monophyletic group.

Examination of diagnostic morphological features for the Russian species (Balushkin & Chereshnev 1982; Chereshnev & Balushkin 1981) in disjunct populations of *D. pectoralis* in Alaska did not find morphological distinction between the populations (Crossman & Rab 1996). However, the karyotypes of fish from North Slope and Yukon River populations examined by Crossman and Rab (1996) show remarkable differences in diploid number and chromosome number variation. A consistent diploid number of 2n=74 for Colville River (North Slope) fish and a wide variation of 2n from 71 to 78 with a proposed diploid number of 2n=76 for Yukon River fish (Beamish et al. 1971; Crossman & Rab 1996). The taxonomic implications of the karyotype differences

observed between populations of *D. pectoralis* in Alaska are either that they are indeed different species, or that the observed variability is part of normal intraspecific variability (Crossman & Rab 1996). A small acrocentric pair of chromosomes may be what is causing the observed difference in diploid number between the North Slope and Yukon River populations; to properly evaluate the value of the karyotype differences, it would be necessary to more broadly sample karyotypes variability among populations of *Dallia* (Crossman & Rab 1996). The same individuals examined by Crossman and Rab (1996) from Galena and the Colville River were examined in this study, allowing a direct comparison of karyotype and mtDNA data.

The mtDNA results do not show North Slope *Dallia* as very divergent and they are nested within the clade of Coastal mitochondrial haplotypes. However, the haplotypes from North Slope *Dallia* are monophyletic within that larger clade. The low diversity of mtDNA in North Slope fish supports the possibility of a founding event. Both the karyological data and our mtDNA indicate substantially different evolutionary histories and long term isolation of *Dallia* populations across the current biogeographical division of the Brooks Range as well as separation along the line of North and South flowing rivers extending across the Bering land bridge into eastern Russia that occurred during the Quaternary glaciations (Lindsey & McPhail 1986).

Both our study and the work of Crossman and Rab (1996) have produced several lines of evidence pointing to the isolation of North Slope *Dallia*. This isolation brings forth questions about taxonomic recognition of North Slope *Dallia* populations.

However, the shallow degree of mitochondrial divergence and absence of nuclear genome data do not permit a proper evaluation of those questions at this time.

Future Research

The analysis of *Dallia* mtDNA has shown very clear phylogeographic patterns, and suggests that within Beringia multiple aquatic refugia existed during the Pleistocene glaciations. The degree of connectivity between Beringian refugia is not clearly elucidated with this mitochondrial dataset. One of these aquatic Beringian refugia appears to be comprised largely of the Bering land bridge itself. How conducive the Bering land bridge was to freshwater fish migration is important in understanding how paleoclimatic instability in the Pleistocene contributed to intercontinental exchange of aquatic organisms. One described species of *Dallia*, *D. delicatissima* was not examined here. The results of this study along with Crossman and Rab (1996) highlight the need for further investigation into the taxonomic status of *Dallia* populations.

Acknowledgements

This study was possible through the generous collecting efforts of and donations of tissues from many individuals. We thank Charles Bille, Willie Takumenjak, Peter Abraham, Mark Lisac, Megan Hoffman, Joseph Buckwalter, Trevor Haynes and his sampling crew, Brendan Scanlon, Dona Eidam, and Frank von Hippel for their assistance in sample collection. The Royal Ontario Museum and University of Washington Burke Museum graciously granted tissue samples from their archives. Tuuli Makinen provided an enormous amount of laboratory and technical assistance. We would also like to extend many thanks to the UAF Spring 2010 Scientific Writing class for their helpful reviews of

the manuscript. Matthew Campbell wishes to recognize his graduate committee member Matt Olson and co-chair Naoki Takebayashi for their assistance throughout his graduate program. Graduate support was provided by the UAF Department of Biology and Wildlife and two graduate research fellowships from Alaska EPSCoR.

References

Alekseyev, S. S., R. Bajno, N. V. Gordeeva, J. D. Reist, M. Power, A. F. Kirillov, V. P. Samusenok, and A. N. Matveev. 2009. Phylogeography and sympatric differentiation of the Arctic charr *Salvelinus alpinus* (L.) complex in Siberia as revealed by mtDNA sequence analysis. *Journal of Fish Biology*, 75, 368-392.

Andreev, A. V. 2004. *Wetlands in Russia, Volume 4: Wetlands in Northeastern Russia.*, Wetlands International-Russia Programme.

Armstrong, R. H. 1994. Alaska blackfish. Alaska Department of Fish and Game Species Profile.

Aspinwall, N. 1965. *Spawning characteristics and early life history of the Alaskan blackfish, Dallia pectoralis, Bean.* M.S. Seattle: University of Washington.

Balushkin, A. V. and I. A. Chereshnev. 1982. Systematics of the genus *Dallia* (Umbridae, Esociformes). [In Russian.] *Trudy Zoologicheskogo Instituta Akademii Nauk SSSR*, 114, pp.36-56.

Bandelt, H. J., P. Forster, and A. Rohl. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16, pp.37-48.

Beamish, R. J., M. J. Merrilees,and E. J. Crossman. 1971. Karyotypes and DNA values for members of the suborder Esocoidei (Osteichthyes: Salmoniformes). *Chromosoma*, 34, pp.436-447.

Benson, D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and D. L. Wheeler. 2005. GenBank. *Nucleic Acids Research*, 33(Database), pp.D34-D38.

Bernatchez, L., and C. C. Wilson. 1998. Comparative phylogeography of Nearctic and Palearctic fishes. *Molecular Ecology*, 7, pp.431-452.

Bernatchez, L., and J. J. Dodson. 1990. Allopatric origin of sympatric populations of lake whitefish (*Coregonus clupeaformis*) as revealed by mitochondrial-DNA restriction analysis. *Evolution*, 44, pp.1263-1271.

Bernatchez, L., and J. J. Dodson. 1991. Phylogeographic structure in mitochondrial DNA of the lake whitefish (*Coregonus clupeaformis*) and its relation to Pleistocene glaciations. *Evolution*, 45, pp.1016-1035.

Blackett, R. F. 1962. Some phases in the life history of the Alaskan blackfish, *Dallia pectoralis*. *Copeia*, 1962, pp.124-130.

Bodaly, R.A., J. W. Clayton, C. C. Lindsey, and J. Vuorinen. 1992. Evolution of lake whitefish (*Coregonus clupeaformis*) in North America during the Pleistocene: genetic differentiation between sympatric populations. *Canadian Journal of Fisheries and Aquatic Sciences*, 49, pp.769-779.

Brunner, P. C., M. R. Douglas, A. Osinov, C. C. Wilson, and L. Bernatchez. 2001. Holarctic phylogeography of Arctic charr (*Salvelinus alpinus* L.) inferred from mitochondrial DNA sequences. *Evolution*, 55, pp.573-586.

Burridge, C. P., D. Craw, D. C. Jack, T. M. King, and J. M. Waters. 2008. Does fish ecology predict dispersal across a river drainage divide? *Evolution*, 62, pp.1484-1499.

Cavender, T. 1969. An Oligocene mudminnow (family Umbridae) from Oregon with remarks on relationships with the Esocoidei. *Occasional Papers of the Museum of Zoology. Number 660*.

Chereshnev, I. A. and A. V. Balushkin. 1981. A new species of blackfish, *Dallia admirabilis sp. n.* (Umbridae, Esociformes), from the Amguema River basin (Arctic Chukotka). *Journal of Ichthyology*, 20, pp.25-30.

Crawford, R. H. 1974. Structure of an air-breathing organ and the swim bladder in the Alaska blackfish, *Dallia pectoralis* Bean. *Canadian Journal of Zoology*, 52, pp.1221-1225.

Crossman, E. J., and P. Rab. 1996. Chromosome-banding study of the Alaska blackfish, *Dallia pectoralis* (Euteleostei: Esocae), with implications for karyotype evolution and relationship of esocoid fishes. *Canadian Journal of Zoology*, 74, pp.147-156.

Drummond, A. J., G. K. Nicholls, A. G. Rodrigo, and W. Solomon. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics*, 161, pp.1307-1320.

Drummond, A. J., A. Rambaut, B. Shapiro, and O. G. Pybus. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, 22, pp.1185 - 1192.

Dupanloup, I., S. Schneider, and L. Excoffier. 2002. A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, 11(12), pp.2571-2581.

Elmer, K. R., J. K. J. Van Houdt, A. Meyer, and F. A. M. Volckaert. 2008. Population genetic structure of North American burbot (*Lota lota maculosa*) across the Nearctic and at its contact zone with Eurasian burbot (*Lota lota lota*). *Canadian Journal of Fisheries & Aquatic Sciences*, 65, pp.2412-2426.

Gudkov, P. K. 1998. Bering Sea *Dallia pectoralis* in the Chukchi Peninsula. *Journal of Ichthyology*, 38, pp.199-203.

Harington, C. R. 1978. Quaternary vertebrate faunas of Canada and Alaska and their suggested chronological sequence. *Syllogeus*, 15, pp.1-105.

Harris, L. N., and E. B. Taylor. 2010. Pleistocene glaciations and contemporary genetic diversity in a Beringian fish, the broad whitefish, *Coregonus nasus* (Pallas): inferences from microsatellite DNA variation. *Journal of Evolutionary Biology*, 23, pp.72-86.

Hewitt, G. 2000. The genetic legacy of the Quaternary ice ages. *Nature*, 405(6789), pp.907-913.

Hopkins, D. M. 1972. The paleogeography and climatic history of Beringia during late Cenozoic time. *Inter-nord*, 12, pp.121-150.

Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17, pp.754 -755.

Jazin, E., H. Soodyall, P. Jalonen, E. Lindholm, M. Stoneking, and U. Gyllensten. 1998. Mitochondrial mutation rate revisited: hot spots and polymorphism. *Nature Genetics*, 18, pp.109-110.

Jordan, D. S., and B. W. Evermann. 1896. The fishes of North and Middle America: a descriptive catalogue of the species of fish-like vertebrates found in the waters of North America, north of the Isthmus of Panama. Part I. *Bulletin of the United States National Museum*, 47, pp.1-1240.

Librado, P., and J. Rozas. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25, pp.1451-1452.

Lindsey, C. C., and J. D. McPhail. 1986. Zoogeography of fishes of the Yukon and McKenzie Basins. In C. H. Hocutt & E. O. Wiley, eds. *Zoogeography of the freshwater fishes of North America*. New York: Wiley Interscience, pp. 639-674.

Lu, G., D. J. Basley, and L. Bernatchez. 2001. Contrasting patterns of mitochondrial DNA and microsatellite introgressive hybridization between lineages of lake whitefish (*Coregonus clupeaformis*); relevance for speciation. *Molecular Ecology*, 10, pp.965-985.

Maddison, W. P. and D. R. Maddison. 2009. *Mesquite: a modular system for evolutionary analysis*, Available at: http://mesquiteproject.org.

Marko, P. B., J. M. Hoffman, S. A. Emme, T. M. McGovern, C. C. Keever, and L. N. Cox. 2010. The 'Expansion-Contraction' model of Pleistocene biogeography: rocky shores suffer a sea change? *Molecular Ecology*, 19, pp.146-169.

McAllister, D. E., S. P. Platania, F. W. Schueler, M. E. Baldwin, and D. D. Lee. 1986. Ichthyofaunal patterns on a geographic grid. In C. H. Hocutt & E. O. Wiley, eds. *Zoogeography of North American Freshwater Fishes*. New York: Wiley Interscience, pp. 17-51.

Mecklenburg, C. W., T. A. Mecklenburg, and L. K. Thorsteinson. 2002. *Fishes of Alaska*, Bethesda, Maryland: American Fisheries Society.

Meldrim, J. W. 1968. *The ecological zoogeography of the Olympic mudminnow,* Novumbra hubbsi *Schultz.* Ph. D. Seattle: University of Washington.

Morrow, J. E. 1980. *The freshwater fishes of Alaska.*, Anchorage: Alaska Northwest Publishing Company.

Nabholz, B., S. Glemin, and N. Galtier. 2008. Strong variations of mitochondrial mutation rate across mammals--the Longevity Hypothesis. *Molecular Biology and Evolution*, 25, pp.120-130.

Nelson, J. S. 2006. *Fishes of the World* 4th ed., Hoboken, New Jersey: John Wiley & Sons, Inc.

Olsen, J., P. Crane, B, Flannery, K. Dunmall, W. Templin, and J. Wenburg. 2011. Comparative landscape genetic analysis of three Pacific salmon species from subarctic North America. *Conservation Genetics*, 12, pp.223-241.

Ostdiek, J. L., and M. N. Roland. 1959. Studies on the Alaskan blackfish *Dallia pectoralis* I. Habitat, size and stomach analyses. *American Midland Naturalist*, 61, pp.218-229.

Phillips, R.B., L. I. Gudex, K. M. Westrich, and A. L. DeCicco. 1999. Combined phylogenetic analysis of ribosomal ITS1 sequences and new chromosome data supports three subgroups of Dolly Varden char (*Salvelinus malma*). *Canadian Journal of Fisheries and Aquatic Sciences*, 56, pp.1504-1511.

Pielou, E. C. 1991. *After the Ice Age: the return of life to glaciated North America*, Chicago: University of Chicago Press.

Posada, D. and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics*, 14, pp.817 -818.

Rambaut, A., and A. J. Drummond. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, 7, p.214.

Rogers, A. R., and H. Harpending. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, 9, pp.552-569.

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19, pp.1572 -1574.

Seeb, L. W., and P. A. Crane. 1999. High genetic heterogeneity in chum salmon in Western Alaska, the contact zone between Northern and Southern lineages. *Transactions of the American Fisheries Society*, 128, pp.58-87.

Seutin, G., B. N. White, and P. T. Boag. 1991. Preservation of avian blood and tissue samples for DNA analyses. *Canadian Journal of Zoology.*, 69, pp.82-90.

Smith, S. A. and C. W. Dunn. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics*, 24, pp.715-716.

Stamford, M., and E. B. Taylor. 2004. Phylogeographical lineages of Arctic grayling (*Thymallus arcticus*) in North America: divergence, origins and affinities with Eurasian *Thymallus*. *Molecular Ecology*, 13, pp.1533-1549.

Turgeon, J., and L. Bernatchez. 2001. Mitochondrial DNA phylogeography of lake cisco (*Coregonus artedi*): evidence supporting extensive secondary contacts between two glacial races. *Molecular Ecology*, 10, pp.987-1001.

Ultsch, G. R. 1989. Ecology and physiology of hibernation and overwintering among freshwater fishes, turtles, and snakes. *Biological Reviews*, 64, pp.435-515.

Van Houdt, J. K. J., L. De Cleyn, A. Perretti, and F. A. M. Volckaert. 2005. A mitogenic view on the evolutionary history of the Holarctic freshwater gadoid, burbot (*Lota lota*). *Molecular Ecology*, 14, pp.2445-2457.

Wilson, C. C. and P. D. N. Hebert. 1998. Phylogeography and postglacial dispersal of lake trout (*Salvelinus namaycush*) in North America. *Canadian Journal of Fisheries and Aquatic Sciences*, 55, pp.1010-1024.

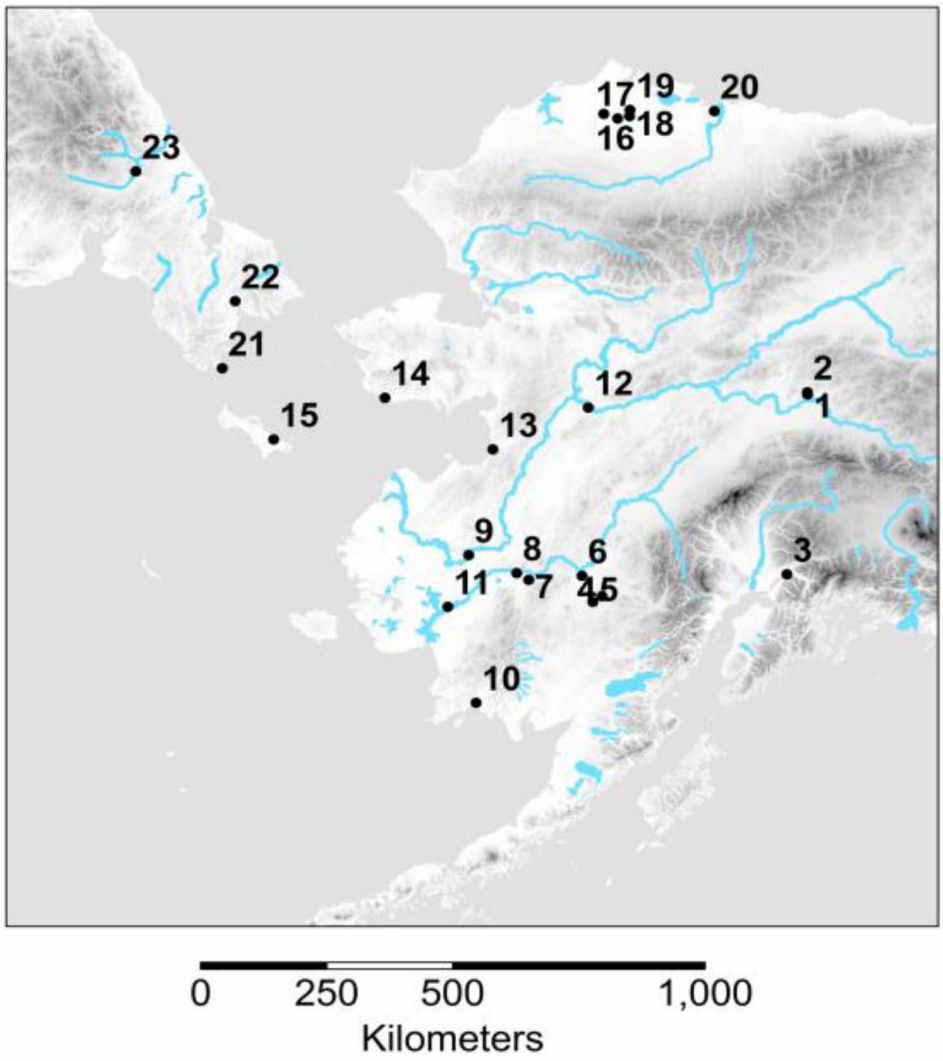

Fig. 1. Map of sample locations of *Dallia* from Alaska, USA (sample locations 1-20) and Chukotka, Russia (sample locations 21-23) used in this study. Coordinates for sample locations are listed in Table 1.

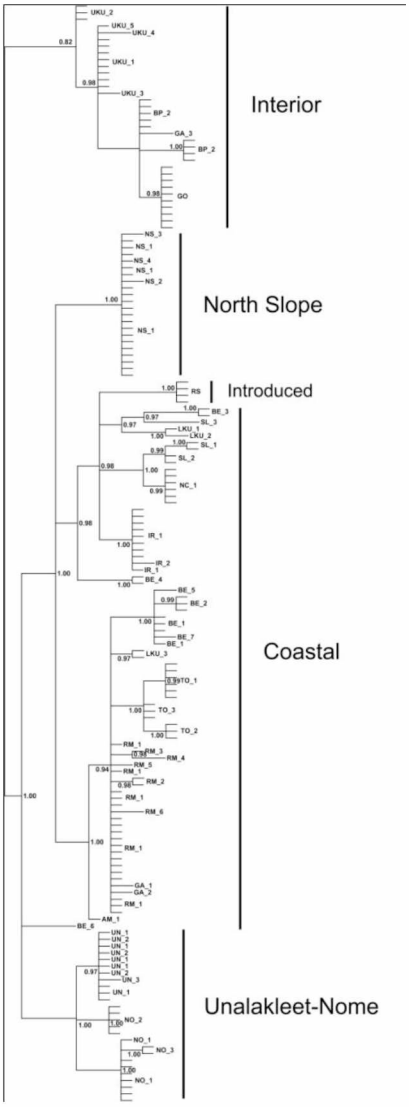

Fig. 2. Bayesian tree of concatenated mtDNA sequences from 168 *Dallia* from Alaska, USA and Chukotka, Russia. Tree was constructed under the general time reversible (GTR) model with a gamma distributed rate variation in four categories (Γ). Bayesian posterior probability values for each clade are included on the tree. Haplotype naming follows Table 2. The four major mitochondrial groups and introduced population from Wasilla, Alaska are identified.

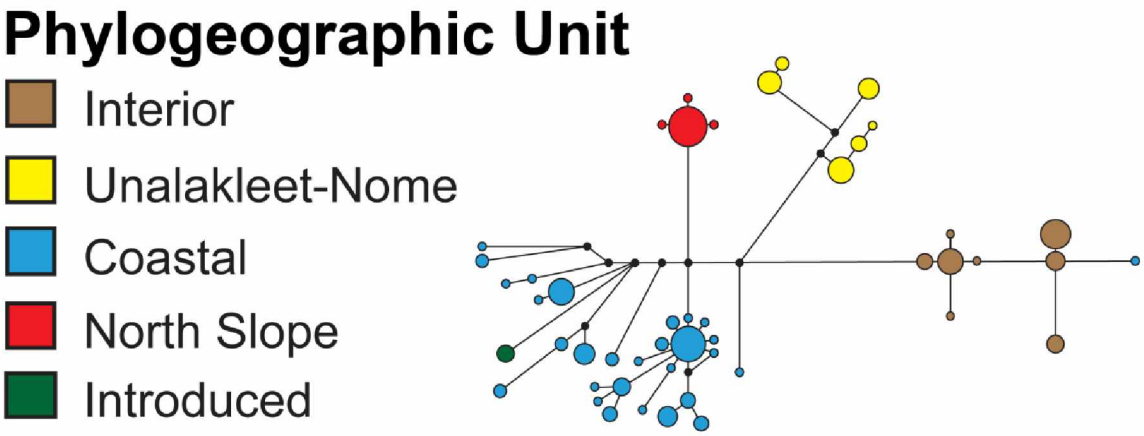

Fig. 3. Median joining network based on *Dallia* concatenated mtDNA sequence data from 168 specimens from Alaska, USA and Chukotka, Russia. Network is color-coded by sampling region: Interior, brown; Unalakleet-Nome, Yellow; Coastal, blue; North Slope, red; Introduced, green.

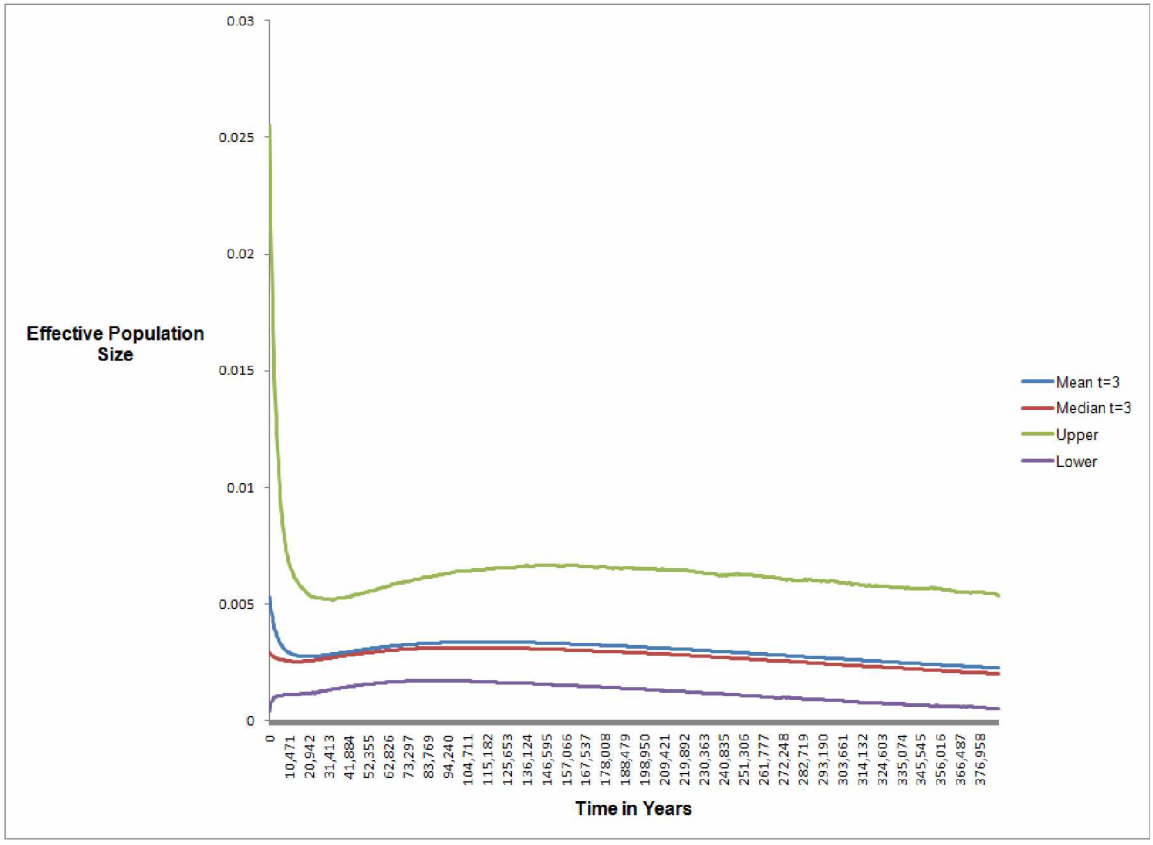

Fig. 4. Bayesian Skyline Plot for the Coastal mitochondrial lineage using concatenated mtDNA from 79 *Dallia* from Alaska, USA and Chukotka, Russia. The x-axis represents time to the lower bound of the 95% highest posterior density interval, the y-axis is effective population size (N_e). Upper and lower estimates of N_e are included, with mean and median values of N_e calculated for a generation time of three years.

10	Togiak	59.0546	-160.3962	12
11	Bethel	60.7904	-161.7799	16
12	Galena	64.7167	-157.0000	12
13	Unalakleet	63.8099	-160.7590	11
14	Nome	64.5061	-165.4305	15
15	St. Lawrence Island	63.3451	-169.4893	5
16	North Slope	70.2768	-156.9182	6
17	North Slope	70.1981	-156.1973	2
18	North Slope	70.2528	-155.5849	3
19	North Slope	70.3683	-155.5697	4
20	Colville River	70.3333	-151.2000	6
21	Novoe Chaplino	64.4085	-172.2590	10
22	Ievineem River	65.6808	-172.5542	10
23	Amguema Basin	67.4346	-178.6985	8

:ation number from

phic coordinates of

1 2008-2010, from

Site Number	Location	Latitude	Longitude	n
1	Fairbanks	64.8692	-147.8254	10
2	Fairbanks	64.9117	-147.8288	10
3	Wasilla	61.5374	-149.2550	5
4	Kuskokwim Basin	61.1945	-156.1535	4
5	Kuskokwim Basin	61.0812	-156.4840	9
6	Kuskokwim Basin	61.5597	-156.9341	1
7	Kuskokwim Basin	61.4300	-158.9114	2
8	Kuskokwim Basin	61.5406	-159.3765	1
9	Russian Mission	61.7952	-161.2443	15
10	Togiak	59.0546	-160.3962	12
11	Bethel	60.7904	-161.7799	16
12	Galena	64.7167	-157.0000	12
13	Unalakleet	63.8099	-160.7590	11
14	Nome	64.5061	-165.4305	15
15	St. Lawrence Island	63.3451	-169.4893	5
16	North Slope	70.2768	-156.9182	6
17	North Slope	70.1981	-156.1973	2
18	North Slope	70.2528	-155.5849	3
19	North Slope	70.3683	-155.5697	4
20	Colville River	70.3333	-151.2000	6
21	Novoe Chaplino	64.4085	-172.2590	10
22	Ievineem River	65.6808	-172.5542	10
23	Amguema Basin	67.4346	-178.6985	8

Table 2. List of concatenated COI and control region mtDNA haplotypes of *Dallia* from Alaska, USA and Chukotka, Russian Federation. The number of fish sequenced with the haplotype is included beneath the sample location (Fig. 1; Table 1) from where the fish was collected.

	Sample location																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
BP_1	5																						
BP_2	4																						
GO		10																					
RS			4																				
UKU_1				1	7																		
UKU_2				3																			
UKU_3					1																		
UKU_4					1																		
UKU_5						1																	
LKU_1							1																
LKU_2							1																
LKU_3								1			1												
RM_1									9			9											
RM_2									2														
RM_3									1														
RM_4									1														
RM_5									1														
RM_6									1														
TO_1										6													
TO_2										3													
TO_3										3													
BE_1											4												
BE_2											3												
BE_3											2												
BE_4											2												
BE_5											1												
BE_6											1												
BE_7											1												
GA_1												1											
GA_2												1											
GA_3												1											
UN_1													7										
UN_2													3										
UN_3													1										
NO_1														8									
NO_2														5									
NO_3														2									
SL_1															2								
SL_2															2								
SL_3															1								
NS_1																4	2	2	4	7			
NS_2																1							
NS_3																1							
NS_4																		1					
NC_1																					6		
IR_1																						6	3
IR_2																						1	
AM_1																							1

Table 3. Fixation indices from SAMOVA for different *K*. Results of the proportion of total genetic variance due to differences between groups (F_{ct}), the proportion of genetic variance due to differences between populations within each group ($F_{sc,}$), and the genetic variance due to differences between populations (F_{st}) for different K in SAMOVA 1.0. All calculations are from a 164 individual, 1,230 base pair COI and control region mitochondrial alignment of *Dallia* from 22 naturally occurring populations.

	K				
	2	3	4	5	6
F_{ct}	0.48	0.56	0.63	0.68	0.71
F_{sc}	0.80	0.74	0.67	0.57	0.52
F_{st}	0.90	0.88	0.88	0.86	0.86

Table 4. Mismatch analyses of mtDNA sequence data. Raggedness statistic (r) from comparison to stable population size, estimates of intrapopulation coalescent time (τ), and upper and lower bounds of T in millions of years (myr) from mismatch analyses conducted on concatenated mtDNA sequence data from 164 *Dallia* individuals sequenced from Alaska, USA and Chukotka, Russian Federation 2008-2010.

Phylogeographic Unit	Raggedness Statistic (r)	τ	T (myr) Upper	Lower
All	**0.004**	**7.35**	**1.84**	**1.23**
Interior	0.099	1.73	0.43	0.29
Unalakleet-Nome	0.107	1.75	0.44	0.29
North Slope	0.450	0.18	0.05	0.03
Coastal	0.013	3.35	0.84	0.56

Chapter 3: Population genetics of *Dallia* in Beringia[2]

ABSTRACT

Background

Pleistocene climatic instability had profound and diverse effects on the distribution and abundance of Arctic organisms resulting in variable phylogeographic patterns among the species of extant Arctic faunas. To understand the effects of paleoclimate on Beringian freshwater fishes, we examined *Dallia* (blackfish: Esociformes: Esocidae). The genus *Dallia* groups small, cold- and hypoxia-tolerant, poorly dispersing fishes restricted entirely in distribution to Beringia. We examined the pattern of genetic diversity to identify distinct refugia within the greater Beringian refugium as well as the roles of the Bering land bridge, Brooks Range, and large rivers within Beringia on the biogeography of *Dallia*. Our analyses were based on DNA sequence data from two nuclear gene introns (S7 and RAG1) and two mitochondrial genome fragments from nineteen sampling locations. These data were examined in clustering and coalescent analyses.

Results

We identified up to five distinct groups of sampled populations of *Dallia*. The clustering and ordination analyses implemented in STRUCTURE and DAPC showed mostly concordant groupings with a high degree of differentiation among groups. The

[2]Campbell, M.A, Takebayashi, N., and Lopez, J.A.L. Population genetics of *Dallia* in Beringia. Prepared for submission to BMC Evolutionary Biology.

groups of sampled populations correspond to geographic areas divided by likely biogeographic barriers. Estimates of sequence diversity (θ) are highest in the Yukon and Kuskokwim River Drainages near the Bering Sea. We found asymmetric estimates of migration rates between populations. The isolation of populations in the North Slope of Alaska was supported by very low estimated migration rates in both directions. Divergence time estimates among populations predate the end of the last ice age.

Conclusions

The estimated divergence time for the groups predates the end of the last glacial period, therefore supporting survival of populations of *Dallia* in multiple refugia throughout Beringia during the Pleistocene. Historic and contemporary isolation across the Brooks Range was identified. Historical migration across the Bering Land bridge resulted in similarities in the genetic composition of West Beringian and Western Coastal Alaska *Dallia* populations.

BACKGROUND

Climatic instability during the Pleistocene strongly influenced the distribution, composition, and genetic diversity of arctic organisms [1]. To date, the study of Holarctic phylogeography has relied heavily on the use of mitochondrial DNA (mtDNA) to investigate the effects of paleoclimatic instability, with mammals receiving most of the

focus [2, 3]. Studies of Arctic phylogeography have identified the biogeographic region known as Beringia as an important glacial refugium for organisms during the Pleistocene glaciations. Beringia, with a unique assemblage of flora and fauna occupies the area adjacent to the former Bering land bridge in both North America and Eurasia [4].

The role of Beringia as an aquatic refugium has been documented in several freshwater fish species [5-11]; however, at least three points are worth highlighting regarding studies of holarctic fish phylogeography. First, the amount of phylogeographic concordance between fishes from glaciated areas is low [12]. That is, due to differential survival and dispersal of genotypes among species the phylogeographic patterns observed differ. Second, the broad geographic scale of holarctic phylogeographic studies results in the sampling effort within Beringia being low. In these studies, the sampling effort has focused on obtaining samples across a species' entire range, which typically is very large in the case of holarctic freshwater fish species. Beringia may be a small component of an organism's range or peripheral to the main distribution of the organism, and therefore is not intensively sampled. Finally, the heavy reliance on mitochondrial data for studies of holarctic freshwater fish phylogeography limits the generality of findings. While mitochondrial DNA offers clear practical advantages in studies of this type, it is maternally inherited as a single non-recombining locus.

Unique among freshwater fishes, *Dallia* (Blackfish; Esociformes: Esocidae) is confined to Beringia and contains up to three species [13]. Study of *Dallia* mtDNA [14]

has identified several potential refugia within the greater Beringia refugium [15]. To better understand the effects of paleoclimate on Beringian freshwater fauna, we further examined the pattern of genetic variability in *Dallia.* We examined DNA sequence variation at mitochondrial and nuclear loci from specimens sampled from across the geographic range of the genus in Alaska, and from three locations in eastern Russia including the type locality of *D. admirabilis*. We investigated the demographic history of *Dallia* with multilocus coalescent methods to estimate genealogical relationships between and within Russian and Alaskan populations of *Dallia*.

In this study we ask the following questions: 1) What are the genetic relationships among sampled locations of *Dallia*? 2) How many Beringian glacial refugia are compatible with observed levels and distribution of genetic variability? 3) Were *Dallia* able to traverse the Bering land bridge, if so, how much and in what direction? Additionally, we provide a genetic perspective on the taxonomic status of different *Dallia* populations within our study.

METHODS

Sample collection

Samples of *D. pectoralis* from localities broadly representing the species range on the Alaska mainland, Saint Lawrence Island, and from two locations on the Chukotka Peninsula were examined (Fig. 1). Samples include populations of *D. pectoralis* from the North Slope of Alaska and fish from the type locality of *D. admirabilis* in Russia to

further refine taxonomic questions in this genus. A total of 188 individuals collected between October 2008 and June 2010 were utilized in this study. Achived tissue samples of Russian *D. pectoralis* and *D. admirabilis* from Siberia from the fish collection at the Burke Museum of Natural History and Culture, University of Washington, Seattle, USA were included in the study (UW 041669, UW 041670, and UW 041671). The samples examined include fishes used in a karyotype and morphology study [16] and were received from the Royal Ontario Museum, Toronto, Canada (TO2653M, TO2654H, TO2655H, TO2656M, TO2657M, TO2658M, TO2659M, TO2660M, TO2689M, TO2690M, TO2690M, TO2691H, TO4305M, TO2692M, TO2693H, TO2696M, TO2697M, TO2699L, TO2700M). DNA was extracted from fin clips if available, but also gill arches and muscle tissue. The procedure is described in the second chapter of this thesis [14].

Sequence data

Two nuclear introns were sequenced from geographically widespread individuals for this study: The second intron of the recombination activating gene I (RAG1) and the first intron of the S7 ribosomal protein.

Primers for the second intron of RAG1 (RAG1 I2) were developed for esociform fishes with the flanking regions of the RAG1 exons on either side of the intron as primer binding sites [17]. We compared sequenced products to known esociform RAG1 I2 intron sequences on GenBank with BLAST to verify amplification of the correct

genomic region. The primer set is forward '5-GAA GTG GAA GCG ATG ATG CAA GGT-3' and reverse 5'-GGC TRC AGC TCA GGA ATG TGT TGA C-3'. PCR conditions for RAG1 I2 were 1X ProMega GoTaq Flexi reaction buffer, 0.2mM dNTP's, 2mM $MgCl_2$, 0.4μM forward primer, 0.4μM reverse primer, 0.025U/μL GoTaq Flexi Taq polymerase, and 1μL template of variable concentrations. For this primer set, the thermocycler was programmed to an initial denaturing of 94°C for 2 min, followed by 30 cycles of denaturing at 94°C for 20 s, annealing at 60°C for 25 s, and an extension step at 72°C for 50 s with a final extension step of 72°C for 2 min. Several degraded samples required us to increase the number of PCR cycles to 40 to generate sufficient product for sequencing.

General fish primers targeting the first intron of the S7 ribosomal protein (S7 1) were used for amplification and sequencing [18]. Sequenced S7 1 PCR amplicons were most similar to other S7 1 sequences available on GenBank. Reaction conditions for S7 1 PCRs were the same as those used for RAG1 in this study. The thermocycler profile was the following: An initial denaturing at 95°C for 2 min followed by 35 cycles and a final extension. Each cycle comprised a 95°C denaturing for 30 s, a 55°C annealing step for 1 min, and a 72°C extension step for 1 min and 30 s. A final 2 min extension step at 72°C ended the profile. Unpurified PCR products were sequenced using ABI Big-Dye v3.1 chemistry in ABI 3730XL machines by High-Throughput Sequencing Solutions at the University of Washington, Seattle, U.S.A.

Mitochondrial DNA (mtDNA) sequences examined in a previous *Dallia* phylogeography study [14] were included in the following analyses. Fragments of mitochondrial gene fragments, cytochrome oxidase I and the mitochondrial control region were concatenated for each individual since the mitochondrial loci behave as a single non-recombining locus. The *Dallia* mtDNA used in this study are publicly available records on GenBank.

Sequences were aligned using CodonCode Aligner version 3.0.3 [19], and haplotypes of nuclear alleles were determined using PHASE version 2.1 [20, 21] implemented in DnaSP version 5.10 [22]. The latter was also used to estimate basic summary statistics of genetic diversity and to test for evidence of recombination with the four-gamete test [22, 23].

Determination of Population Structure

We used three different methods to evaluate population structure since each method has different strengths and weaknesses. STRUCTURE version 2.3.3 [24, 25] and InStruct [26], are clustering methods that use Bayesian clustering algorithms to infer population substructure and assign each individual to the sub populations. However, one key assumption of STRUCTURE implementation of this clustering analysis is that of Hardy-Weinberg equilibrium within each subpopulation. Therefore, we also examined our data for population structure using InStruct. InStruct, an extension of the STRUCTURE

model, relaxes the assumption of Hardy-Weinberg equilibrium to accommodate the potential effects of inbreeding. We also used a third approach, Discriminant Analysis of Principal Components (DAPC) as implemented in an R package, *adegenet* [27], to identify and characterize genetic structure among populations. Unlike the two clustering methods, DAPC is not based on an underlying population genetics model. DAPC accounts for arbitrary linkage structures among single nucleotide polymorphic sites by transforming the polymorphisms into principal components [28]. The conversion from genetic data into principal components permits the use of generic clustering techniques such as K-means clustering and discriminant analysis.

Unique mitochondrial haplotypes and nuclear alleles were coded as numbers and entered into STRUCTURE version 2.3.3 [24]. Using a burn-in of 100,000 steps followed by a MCMC of 200,000 steps, the likelihood of number of subdivisions (K) within *Dallia* were calculated for $K=\{1,...,8\}$. The same infile that we developed for use with STRUCTURE was used with InStruct. We ran InStruct with two chains, a burn-in of 500,000 steps, a MCMC of 1,000,000 steps, and a thinning interval of 10. We allowed InStruct to find the optimum number of subdivisions within our data set and calculate population inbreeding coefficients. Graphs of the output from both of these programs were made using Distruct [29].

We used an implementation of DAPC in a package for R statistical software [27, 30]. DAPC identifies groups of individuals from among the sample like STRUCTURE and

InStruct, but the underlying methodology is very different. To use DAPC, we converted our sequence data to single nucleotide polymorphisms (SNPs). The SNP data were imported into R as separate diploid and haploid (mtDNA) data and then combined with the *adegenet* [27] package. To perform DAPC we needed to identify population clusters *ex nihilo*, which we did using the *find.clusters* function within *adegenet*. For each *k* of population clusters, *find.clusters* calculates the Bayesian information criterion (BIC) for the corresponding K-means for each *k*. The function *kmeans* is distributed as part of the basic R *stats* package [31]. The resulting plot of BIC scores for K-means of each *k*, aids identification of the most appropriate number of *k* groups. The population clusters identified by *find.clusters* were used as prior assignments in Discriminant Analysis of Principal Components (DAPC), as the function *dapc* in the *adegenet* package [27, 28]. DAPC first performs Principal Component Analysis (PCA, *dudi.pca*) and Discriminant Analysis (DA, *Ida*) utilizing the packages *ade4* [32-34] and *mass* [35]. The clusters identified through these analyses were then used as populations in subsequent coalescent analyses.

Demography and pattern of gene flow

Pairwise comparisons between populations of *Dallia* were performed in Isolation with Migration (IM) [36, 37] to estimate the patterns of connectivity among the populations identified in the analyses described above. Additional mitochondrial sequence data from sampling locations used in quantifying population structure were included if available.

Salient parameters of demographic history, such as population size, divergence time and migration rates can be estimated with an IM comparison. We can also evaluate the effect of biogeographic barriers such as the Bering Sea by comparing migration rates between populations on either side of this barrier. To identify major patterns of migration, separation, and genetic diversity across the range of *Dallia*, we used a series of pairwise IM comparisons

IM uses a Markov Chain Monte Carlo (MCMC) algorithm to simulate genealogies based on a two-population and six-parameter model. IM estimates the divergence time (t) at some time in the past for two populations. Since time t, migration rate in each direction between the two descendent populations are given (m_1 and m_2). The θ for three populations are estimated: the ancestral population, and both descendent populations (θ_A, θ_1, and θ_2.)

Each pairwise comparison between *Dallia* populations consisted of several IM analyses. Initially, prior distributions with wide ranges were specified for the six parameters to determine suitable upper limits for priors in subsequent analyses. Several iterations of assessing convergence and constraining priors were conducted. Final convergence was evaluated by repeated long MCMC searches with differing seed numbers and comparing results. Conversion of t to years required mutation rates to be defined for all loci. We defined sets of upper and lower bound of mutation rates to provide minimum and maximum estimates of divergence. We aligned our RAG1 I2

sequences from *Dallia* to six Esocidae relatives [17], representing the most closely related extant taxa to *Dallia*. The average number of nucleotide substitutions per site between *Dallia* and its relatives, D_{xy}, was calculated in DnaSP version 5.10 [22]. In order to translate the measure of differentiation into rate, we used two different times. At a minimum, esocid fishes were present at the end of the Cretaceous, approximately 65.5 million years ago [38]. However, the divergence between Umbridae and Esocidae is not clearly pinpointed, and the well-studied divergence between Esociformes and Salmoniformes is at least 200 million years ago [39-42]. We decided on an upper bound of 120 million years for the potential divergence between *Dallia* and other esocids.

RESULTS

DNA Sequencing

The total number of individuals sequenced for RAG1 I2 is 70 and S7 1 is 77. We obtained previously determined mtDNA sequence data [14] from 124 individuals (Table 1).

For both nuclear loci, haplotypic phase of heterozygous individuals was resolved with high posterior probabilities (> 0.90) for all but three individuals in the RAG1 I2 alignment and another set of three individuals in the S7 1 intron alignment. We have used the highest probability phase for the following analyses and the robustness of assignment is confirmed by separate iterations of PHASE with different seeds. Ambiguous sites within an individual not having a phase posterior probability > 0.90 are

coded as missing data for use in analyses requiring phase information such as IM and recombination test. For calculating site-frequency spectrum based summary statistics (Table 1), all data including unphased sequences are included. DAPC does not require phase info since it is based on SNP data, therefore SNP data from low posterior probability for phase were included.

Determination of Population Structure

The optimum number of groups of the nineteen sampled populations found by the STRUCTURE algorithm is four (Figs 2 & 3). The four groups correspond to four geographic areas: 1) the Tanana River drainage and Kuskokwim River drainage upstream of the Kuskokwim Mountains (Interior Alaska, sample locations 1, 2, & 4-6); 2) the North Slope of Alaska (sample locations 16-20); 3) Western Coastal Alaska and the Yukon and Kuskokwim River systems downstream of sample locations 1-2 & 4-6 (Coastal Alaska); and 4) Russian and Nome sample locations (21, 22, and 14). STRUCTURE provides estimates of fixation index, F_{st}, from an ancestral population for each of the four groups of sampled populations identified. For the groups listed above, the F_{st} values in order are: 0.47, 0.68, 0.0033, and 0.44. InStruct chooses the optimum number of groups of sampled populations as four (Fig 2). However, individuals are highly admixed at K=4 and the variance of log-likelihood scores for each K become very high. The increased number of parameters required for InStruct (separate inbreeding

coefficient for each *K* populations) likely results in an over-parameterization and flat likelihood surface.

Results from DAPC indicate that there are five separate groups of sampled populations (Figs 2 & 4). Group1 consists of fish from the North Slope (sample locations 16-20). Group 2 contained fish from the Nome sample location (14). Group 3 is composed of the Tanana River and Kuskokwim River upstream of the Kuskokwim Mountains (Interior Alaska, sample locations 1, 2, & 4-6). Group 4 is made up of fish from sample locations in Chukotka, Russia (sample locations 21 & 22). Group 5 contained Western Coastal Alaska sample locations including Yukon and Kuskokwim locations downriver from sample locations 1, 2, & 4-6 (Coastal Alaska, sample locations 7-13). Probability of assignment into all groups is very high for all individuals. Only four individuals are assigned to groups of fish from separate sampling locations. A single fish from sample location 12 is assigned to Group 3 and three fish from the lower Kuskokwim River drainage (sample locations 7 and 11) are placed in Group 4 with fish from Russia. F_{st} estimates were calculated from groups identified in DAPC in a pair-wise fashion and are shown in Table 2.

Demography and Pattern of Gene Flow

Data from 19 sampling locations were grouped into five distinct populations as indicated by DAPC. We omitted the Nome population for use in IM due to small sample

size, and therefore used four populations: Coastal, West Beringia, Interior, and North Slope populations (Table 3). The Coastal population contains fishes from Western Coastal Alaska, the Yukon River, and the Lower Kuskokwim River. West Beringia contains fishes collected from the Chukotka Peninsula of Russia. The Interior population contains fishes from the Upper Kuskowkim River and Tanana River drainage, while the North Slope population consists of fishes collected from the North Slope of Alaska. Nome population fish were collected from the vicinity of Nome, Alaska.

We conducted three pairwise comparisons between the Coastal population and the other three populations excluding Nome due to the small sample size representing that group.

Migration rates

Estimates of migration rates between populations in the IM comparisons were in general asymmetrical and ranged between 0.08 and 2.42 individuals per generation on average since the time of divergence between each population pair (Fig 5). The migration rate from the Coastal population to the other three populations was consistently smaller than the migration rate from the other populations to the coastal populations in the pairwise comparisons.

Divergence times

Divergence time (t), from the IM analyses between populations was estimated for all of the pairwise comparisions (Table 4). The deepest divergence is estimated between the North Slope and Coastal populations, with the shallowest occurring between the Interior and Coastal populations. Estimates of the t parameter range from 3.55 to 12.57. Using a calculated D_{xy} of 0.13483, the mutation rate of RAG1 I2 was estimated with a lower bound of 1.03×10^{-9} and an upper bound of 5.61×10^{-10}. The S7 1 intron is given the same mutation rates since no suitable calibration point was available. The mtDNA is treated as mutating at 1.5×10^{-8} [43]. Conversion to years for all of the lowest 90% highest probability density predates the Last Glacial Maximum (LGM) for both geometric means of per locus mutation rate used, 2.20×10^{-6} and 1.47×10^{-6}.

Estimates of θ

The estimated θ for the Coastal population is consistently much larger than the populations with which it was compared (Table 5). There is no overlap in the 90% highest probability density between the estimate of θ for the Coastal population and the opposing populations in pair-wise comparisons.

DISCUSSION

Determination of Population Structure

Our comparison of three population clustering implementations (STRUCTURE, InStruct, and DAPC) provided consistent evidence of strong population structure across the current geographic range of *Dallia*. Subdivision of *Dallia* within Beringia into up to five different areas was supported by DAPC. Generally, DAPC and STRUCTURE produced divisions among sampling locations in the same way. Only DAPC placed fish from the Nome sampling location in their own group. The lack of support in STRUCTURE for more than four populations is likely a result of haplotype coding instead of SNP coding which was used in DAPC. Nonetheless, all three methods consistently supported a North-South Beringian divide in population structure, which has been observed in other Beringian fishes and hypothesized for *Dallia* [7, 11, 44-46]. Additionally, the five identified groups of sampled populations correspond to major geographic areas and barriers such as the Brooks Range, Bering Sea, the Kuskokwim Mountains and Yukon/Tanana rivers.

Coalescent Analysis

Our estimates of migration patterns appear to be mostly compatible with geographic features, and there is moderate amount of isolation. Migration across or around the Brooks Range is inferred to be very low due to the low estimated migration between the Coastal and North Slope populations, much less in both directions than one individual

per generation. Migration of less than one individual per generation is low enough to allow differentiation under the island model [47]. The Yukon and Kuskokwim Rivers promote asymmetrical migration in a downstream fashion between the Interior and Coastal population. The high rate of downstream migration is compatible with the low dispersal ability expected from *Dallia*. In two cases, the IM model estimates migration to occur across what should be impassable barriers for this species. It is very unlikely that *Dallia* are currently moving across the Brooks Range or the Bering Sea. Since the migration rate is integrated across time since the populations' divergence, the estimated migration rates are most likely the product of historically higher migration in the case of the Brooks Range and Bering Sea.

The large estimates of divergence times between populations of *Dallia* in Beringia from the coalescent analysis clearly suggest that extant *Dallia* populations are not the product of range expansion from a single glacial refugium after the end of the LGM about 12,000 years before present [48]. Due to the inherently large variation in the coalescent process, estimated divergence time is accompanied by a wide range of credibility intervals; therefore the observed divergences are compatible with a diverse set of paleogeographic conditions. Nonetheless, the analysis demonstrates that *Dallia* from Western Alaska and West Beringia were exchanging migrants fairly recently and were not historically isolated over a long time in West Beringia. The Interior Alaska *Dallia* seem to have diverged more recently from the Coastal population, which given that the two populations occur along the Kuksokwim and Yukon Rivers is a likely scenario.

Divergence between North Slope and Coastal populations likely predates the LGM, with potential for long term isolation.

We were unable to generate reliable estimates of ancestral θ in this study. However, it is likely that ancestral population's θ was much larger than our estimates of θ for contemporary populations. The effects of paleoclimatic instability on *Dallia* likely caused a reduction in ancestral θ. Fossil evidence of *Dallia* suggest that prior to the Pleistocene glaciations this genus had a much larger range, extending up to 800 km farther west than its current distribution, and South of the Alaska Range in Alaska where it does not naturally occur today [13, 49, 50]. Furthermore, while post-Pleistocene range expansion is typical for many fish species via pro-glacial lake mediated dispersal [51], it appears that the range of *Dallia* was constricted by paleoclimatic instability. The same pattern of range reduction during the Pleistocene occurred in the Pacific Northwest in *Novumbra*, an ecologically similar relative of *Dallia* [52]. *Novumbra* has one extant species, and while fossil evidence shows the genus was once more widespread in the Pacific Northwest, *Novumbra* currently occurs only in Western Washington State, USA as a result of climatic changes during the Pleistocene [49, 52].

It is very likely that through times of moderate conditions during interglacial periods *Dallia* survived in higher numbers and were able to migrate more effectively due to the generally much wetter conditions [48]. However, changes in climate that reduced

precipitation and cooled the climate could have severely impacted *Dallia* in many parts of its range leading to isolation and reduction in population sizes and genetic variability. The Bering land bridge area on the other hand contained large rivers and associated deltas, low gradient, and wetter habitat [13, 48]. The Coastal population contains sample locations that historically were part of or very close to the hypothetically suitable *Dallia* habitat and is central in the current distribution of the genus. Therefore, the large estimated θ for the Coastal population relative to the other populations suggests that the Bering land bridge and immediately adjacent parts of Alaska in the Yukon-Kuskokwim Delta region historically were and continue to be a center of *Dallia* genetic diversity.

CONCLUSIONS

The data and analyses presented here support the hypothesis that Beringia contained up to five distinct areas where *Dallia* could survive during glacial advances. Broadly defined, these geographic areas include the Bering land bridge and adjacent areas of Russia and Alaska, the vicinity of Nome, Alaska on the southern side of the Seward Peninsula, the Tanana River and Kuskokwim River drainage, and the North Slope of Alaska.

AUTHORS' CONTRIBUTIONS

MC PARTICIPATED IN STUDY DESIGN, GENERATED MOLECULAR DATA, PERFORMED ANALYSES, AND DRAFTED THE MANUSCRIPT. JL PARTICIPATED IN STUDY DESIGN AND CONTRIBUTED TO THE DRAFTING OF THE MANUSCRIPT. NT PARTICIPATED IN STUDY DESIGN, INTERPRETATION OF ANALYSES, AND DRAFTING OF MANUSCRIPT. ALL AUTHORS READ AND APPROVED THE FINAL MANUSCRIPT.

ACKNOWLEDGEMENTS

Matthew Campbell wishes to thank his committee member Dr. Matthew Olson for assistance in the preparation of this manuscript. We made a great deal of use of the UAF Life Science Informatics computational portal: UAF Life Science Informatics as a core research resource is supported by Grant Number RR016466 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH). Matthew Campbell was supported by the University of Alaska Fairbanks and Alaska EPSCoR while generating data and preparing this manuscript.

REFERENCES

1. Hewitt G: **The genetic legacy of the Quaternary ice ages**. *Nature* 2000, **405**:907-913.

2. Shafer ABA, Cullingham CI, Côté SD, Coltman DW: **Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America**. *Mol Ecol* 2010, **19**:4589-4621.

3. Hewitt G: **The structure of biodiversity - insights from molecular phylogeography**. *Front Zool* 2004, **1**:4.

4. Pielou EC: **After the Ice Age: the return of life to glaciated North America**. Chicago: University of Chicago Press; 1991.

5. Harris LN, Taylor EB: **Pleistocene glaciations and contemporary genetic diversity in a Beringian fish, the broad whitefish, *Coregonus nasus* (Pallas): inferences from microsatellite DNA variation**. *J Evol Biol* 2010, **23**:72-86.

6. Bernatchez L, Dodson JJ: **Phylogeographic structure in mitochondrial DNA of the lake whitefish (*Coregonus clupeaformis*) and its relation to Pleistocene glaciations**. *Evolution* 1991, **45**:1016-1035.

7. Brunner PC, Douglas MR, Osinov A, Wilson CC, Bernatchez L: **Holarctic phylogeography of Arctic charr (*Salvelinus alpinus* L.) inferred from mitochondrial DNA sequences.** *Evolution* 2001, **55**:573-586.

8. Wilson CC, Hebert PDN: **Phylogeography and postglacial dispersal of lake trout (*Salvelinus namaycush*) in North America**. *Can J Fish Aquat Sci*1998, **55**:1010-1024.

9. Houdt JKJV, Cleyn LD, Peretti A, Volckaert FAM: **A mitogenic view on the evolutionary history of the Holarctic freshwater gadoid, burbot (*Lota lota*)**. *Mol Ecol* 2005, **14**:2445-2457.

10. Elmer KR, Van Houdt JKJ, Meyer A, Volckaert FAM: **Population genetic structure of North American burbot (*Lota lota maculosa*) across the Nearctic and at its contact zone with Eurasian burbot (*Lota lota lota*).** *Can J Fish Aquat Sci* 2008, **65**:2412-2426.

11. Stamford M, Taylor EB: **Phylogeographical lineages of Arctic grayling (*Thymallus arcticus*) in North America: divergence, origins and affinities with Eurasian *Thymallus***. *Mol Ecol* 2004, **13**:1533-1549.

12. Bernatchez L, Wilson CC: **Comparative phylogeography of Nearctic and Palearctic fishes**. *Mol Ecol* 1998, **7**:431-452.

13. Lindsey CC, McPhail JD: **Zoogeography of fishes of the Yukon and McKenzie Basins**. In **Zoogeography of the freshwater fishes of North America**. edited by Hocutt CH, Wiley EO New York: Wiley Interscience; 1986:639-674.

14. Campbell MA, Lopez JA: **Mitochondrial phylogeography of a Beringian endemic.** In **Mitochondrial phylogeography and population genetics of a Beringian endemic: *Dallia* (Esociformes: Teleostei). Campbell MA, 2011**.

15. Gómez A, Lunt D: **Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula**. In **Phylogeography of southern European refugia**. Dodrecht: Springer Netherlands; 2007:155-188.

16. Crossman EJ, Rab P: **Chromosome-banding study of the Alaska blackfish, *Dallia pectoralis* (Euteleostei: Esocae), with implications for karyotype evolution and relationship of esocoid fishes.** *Can J Zool* 1996, **74**:147-156.

17. Grande T, Laten H, López JA: **Phylogenetic relationships of extant esocid species (Teleostei: Salmoniformes) based on morphological and molecular characters**. *Copeia* 2004, **2004**:743-757.

18. Chow S, Hazama K: **Universal PCR primers for S7 ribosomal protein gene introns in fish.** *Mol Ecol* 1998, **7**:1255-1256.

19. **Sequence assembly and alignment software - CodonCode** [http://www.codoncode.com/].

20. Stephens M, Smith NJ, Donnelly P: **A new statistical method for haplotype reconstruction from population data**. *Am J Hum Genet* 2001, **68**:978-989.

21. Stephens M, Donnelly P: **A comparison of Bayesian methods for haplotype reconstruction from population genotype data**. *Am J Hum Genet* 2003, **73**:1162-1169.

22. Librado P, Rozas J: **DnaSP v5: a software for comprehensive analysis of DNA polymorphism data**. *Bioinformatics* 2009, **25**:1451-1452.

23. Hudson RR, Kaplan NL: **Statistical properties of the number of recombination events in the history of a sample of DNA sequences**. *Genetics* 1985, **111**:147-164.

24. Falush D, Stephens M, Pritchard JK: **Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies**. *Genetics* 2003, **164**:1567-1587.

25. Wilson I, Weale M, Balding D: **Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities**. *J Roy Stat Soc A Sta* 2003, **166**:155-188.

26. Gao H, Williamson S, Bustamante CD: **A Markov Chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multi-locus genotype data**. *Genetics* 2007, **176**:1635-1651.

27. Jombart T: **adegenet: a R package for the multivariate analysis of genetic markers**. *Bioinformatics* 2008, **24**:1403 -1405.

28. Jombart T, Devillard S, Balloux F: **Discriminant analysis of principal components: a new method for the analysis of genetically structured populations**. *BMC Genetics* 2010, **11**:94.

29. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW: **Genetic structure of human populations**. *Science* 2002, **298**:2381 - 2385.

30. R Development Core Team: *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2011.

31. **R: the R stats package** [http://stat.ethz.ch/R-manual/R-patched/library/stats/html/00Index.html].

32. Chessel D, Dufour A: **The ade4 package-I- One-table methods.** *R News* 2004, **4**:5-10.

33. Dray S, Dufour A, Chessel D: **The ade4 package- II: Two-table and K-table methods.** *R News* 2007, 7:47-54.

34. Dray S, Dufour A: **The ade4 package: implementing the duality diagram for ecologists.** *J Stat Softw* 2007, **22**:1-20.

35. Venables W, Ripley B: **Modern Applied Statistics with S**. Fourth. New York: Springer; 2002.

36. Nielsen R, Wakeley J: **Distinguishing migration from isolation: a Markov Chain Monte Carlo approach**. *Genetics* 2001, **158**:885-896.

37. Hey J, Nielsen R: **Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura and D. persimilis***. *Genetics* 2004, **167**:747-760.

38. Wilson MVH, Brinkman DB, Neuman AG: **Cretaceous Esocoidei (Teleostei): early radiation of the pikes in North American fresh waters**. *J Pal* 1992, **66**:839-846.

39. Hedges SB, Dudley J, Kumar S: **TimeTree: a public knowledge-base of divergence times among organisms**. *Bioinformatics* 2006, **22**:2971 -2972.

40. Hurley IA, Mueller RL, Dunn KA, Schmidt EJ, Friedman M, Ho RK, Prince VE, Yang Z, Thomas MG, Coates MI: **A new time-scale for ray-finned fish evolution**. *Proc R Soc B* 2007, **274**:489 -498.

41. Inoue JG, Miya M, Venkatesh B, Nishida M: **The mitochondrial genome of Indonesian coelacanth *Latimeria menadoensis* (Sarcopterygii: Coelacanthiformes) and divergence time estimation between the two coelacanths**. *Gene* 2005, **349**:227-235.

42. Peng Z, He S, Wang J, Wang W, Diogo R: **Mitochondrial molecular clocks and the origin of the major Otocephalan clades (Pisces: Teleostei): A new insight**. *Gene* 2006, **370**:113-124.

43. Ho SYW, Phillips MJ, Cooper A, Drummond AJ: **Time dependency of molecular rate estimates and systematic overestimation of recent divergence times**. *Mol Biol Evol* 2005, **22**:1561 -1568.

44. Balushkin AV, Chereshnev IA: **Systematics of the genus *Dallia* (Umbridae, Esociformes).** [In Russian] *Proc. Zool. Inst. Acad. Nauk SSSR* 1982, **114**:36-56.

45. Chereshnev IA, Balushkin AV: **A new species of blackfish, *Dallia admirabilis sp. n.* (Umbridae, Esociformes), from the Amguema River basin (Arctic Chukotka)**. *J. Ichthyol* 1981, **20**:25-30.

46. Seeb LW, Crane PA: **High genetic heterogeneity in chum salmon in Western Alaska, the contact zone between Northern and Southern lineages**. *T Am Fish Soc* 1999, **128**:58-87.

47. Gillespie JH: **Population Genetics: A Concise Guide**. 2nd edition. Baltimore: Johns Hopkins University Press; 2004.

48. Hopkins DM: **The paleogeography and climatic history of Beringia during late Cenozoic time.** *Inter-nord* 1972, **12**:121-150.

49. Cavender T: **An Oligocene mudminnow (family Umbridae) from Oregon with remarks on relationships with the Esocoidei.** *Occas Pap Mus Zool Number 660* 1969.

50. Harington CR: **Quaternary vertebrate faunas of Canada and Alaska and their suggested chronological sequence.** *Syllogeus* 1978, **15**:1-105.

51. McAllister DE, Platania SP, Schueler FW, Baldwin ME, Lee DD: **Ichthyofaunal patterns on a geographic grid.** In *Zoogeography of North American Freshwater Fishes*. edited by Hocutt CH, Wiley EO New York: Wiley Interscience; 1986:17-51.

52. Meldrim JW: **The ecological zoogeography of the Olympic mudminnow, *Novumbra hubbsi* Schultz.** 1968.

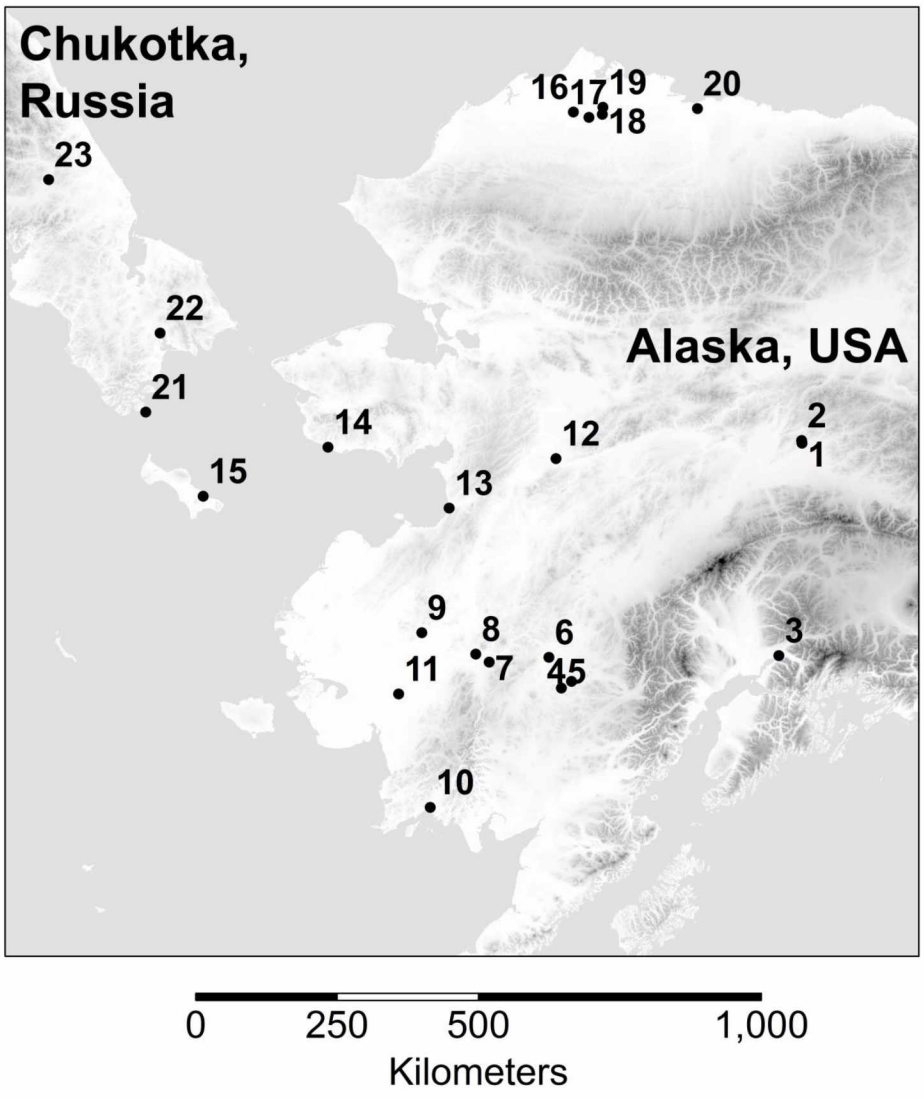

Figure 1 - Sampling locations of Dallia in Beringia.

Distribution of sampling locations used in this study, coordinates available in Chapter 2 of this thesis [14].

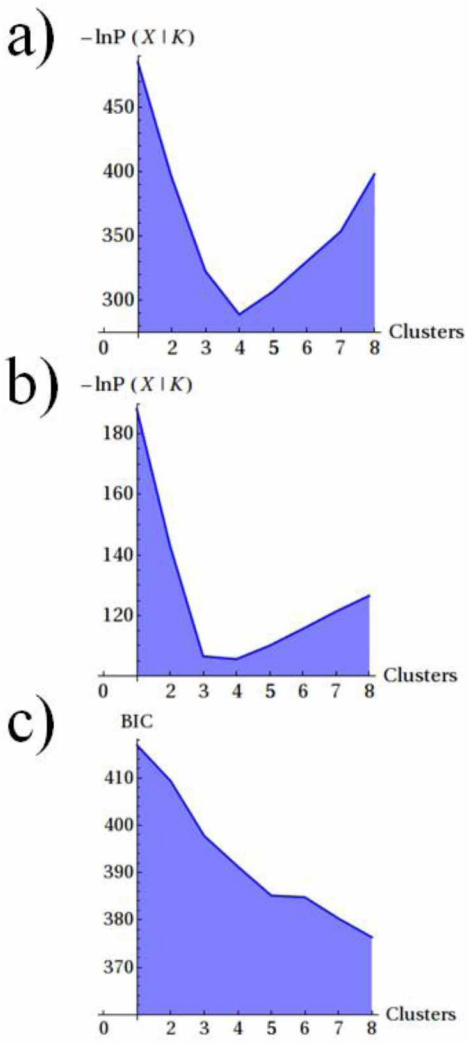

Figure 2 - Data supporting number of groups of sampled populations (K) from STRUCTURE, InStruct, and DAPC.

Negative log likelihood (-lnP (X | K)) or Bayesian Informative Criterion (BIC) support for the number of groups of sampled populations (K) for K={2,... ,8}. Methods of choosing K are shown for STRUCTURE, InStruct, and DAPC.

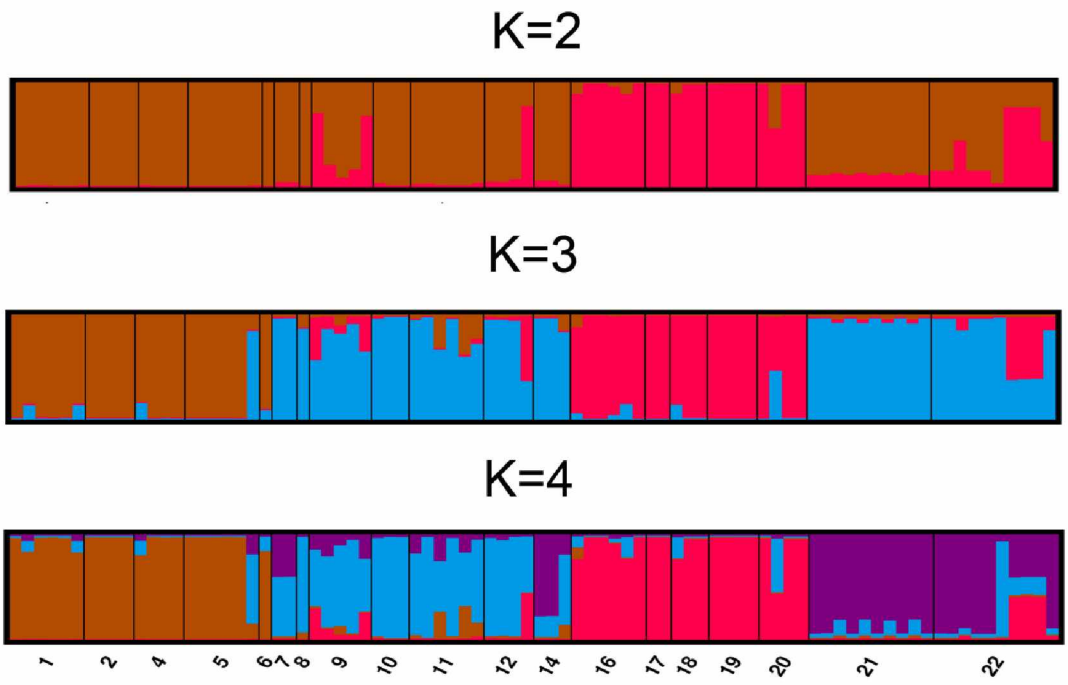

Figure 3 - Comparison of Distruct graphs of STRUCTURE for groups of sampled populations of two through four.

Distruct graphs of STRUCTURE output for number of groups of sampled populations (*K*) for *K*={2,... ,4}. Sampling location number is indicated on the X axis and corresponds to Figure 1.

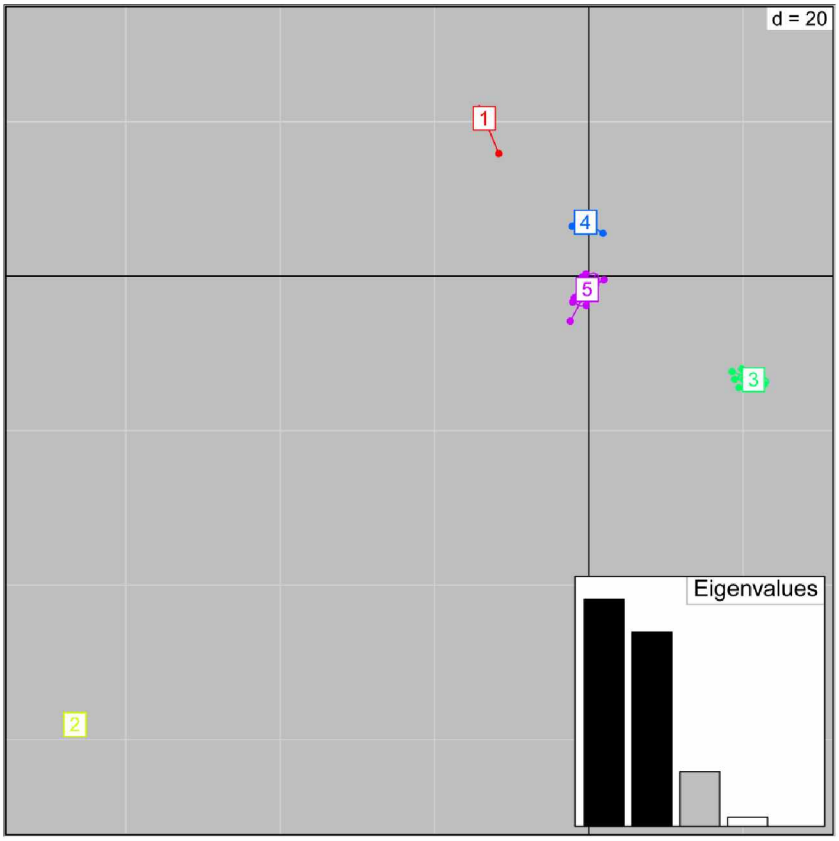

Figure 4 - DAPC scatterplot for K=5.

DAPC scatterplot for *K*=5, depicting the clustering of individuals based on K-means clustering of principal components including eigenvalues representing between to within group variation for linear combinations of principal components. Cluster 1 represents 19 individuals from sample locations 16-20 (North Slope); Cluster 2 has three fish from sample location 16 (Nome); Cluster 3 contains 22 fish from sample locations 1&2, and 4-6 with one individual from sample location 12 (Interior Alaska); Cluster 4 contains 22 fish from sample locations 21 and 22 with three fish each from sample locations 7 and 11 (West Beringia); 17 individuals from sample locations 9-12 are placed in Cluster 5 (Coastal Alaska).

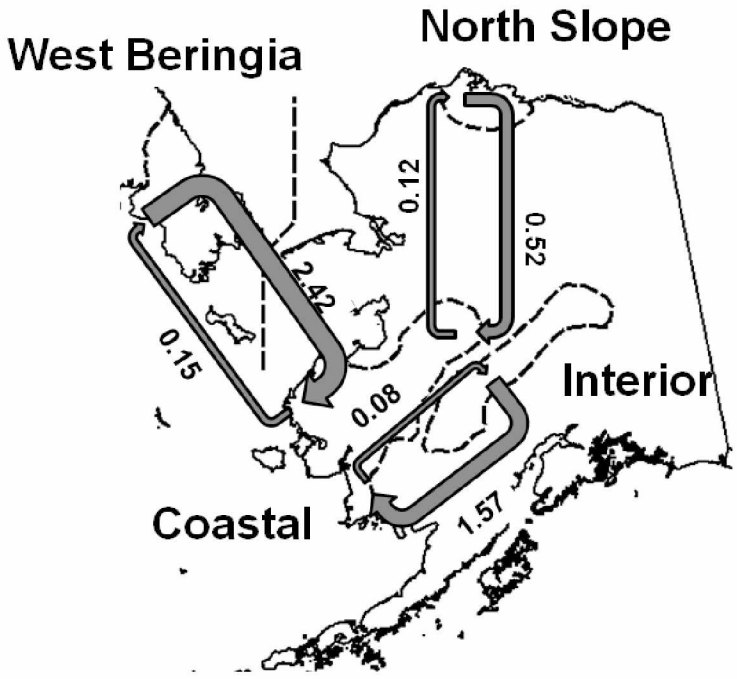

Figure 5 - Migration rates among Beringian Dallia populations.

Migration rates in individuals per generation since the time of divergence (*t*) between populations of *Dallia* estimated by Isolation with Migration analyses. Arrows reflect the magnitude of migration rate.

Table 1 – Summary of alignment length, polymorphism, and tests of neutrality for the mtDNA, RAG1 I2 and S7 1 alignments used in this study.

Descriptions of each of the three alignments in this study, including number of sequences, length, basic polymorphism data, and tests of neutrality. Statistically significant neutrality tests are indicated with an asterisk (*).

Alignment	Number of Sequences	Alignment Length (base pairs)	Variable Sites	Number of Haplotypes	Haplotype Diversity (HD)
mtDNA	124	1230	64	35	0.94
RAG1 I2	128	730	7	5	0.50
S7 1	146	750	7	8	0.50

	Average Pairwise Differences (k)	Nucleotide Diversity (π)	Fu and Li's D	Fu and Li's F	Tajima's D
mtDNA	10.31	0.0084	0.59	0.62	0.42
RAG1 I2	1.78	0.0024	0.27	0.56	0.86
S7 1	0.55	0.00073	1.59	1.75	1.25

Table 2 – DAPC estimates of F_{st} based on pair-wise comparisons of groups of individuals identified.

Each of the five groups of sampled are compared to each other to calculate F_{st}.

Geographic Area	Russia	Interior Alaska	North Slope	Nome
Interior Alaska	0.63	-	-	-
North Slope	0.61	0.38	-	-
Nome	0.85	0.67	0.67	-
Coastal Alaska	0.49	0.36	0.32	0.91

Table 3 - Number of locations and sequences used for each population in IM analyses.

Each population used for IM analyses in this study and the number of sampling locations that compose each population are listed. The number of sequences of each type of genetic data is also summarized.

	Number of:				
Population	**Sampling Locations**	**Individuals**	**mtDNA Sequences**	**RAGI I2 Sequences**	**S7 1 Sequences**
Coastal	**6**	**58**	**58**	**26**	**36**
West Beringia	**2**	**19**	**13**	**28**	**38**
Interior	**5**	**32**	**32**	**40**	**34**
North Slope	**5**	**21**	**21**	**34**	**36**

Table 4 - Estimates of divergence time (t) from the three Isolation with Migration comparisons.

Estimate of divergence time (*t*) in represented in the mean of the *t* parameter and estimated in years for the mean of *t* and 90% highest probability density estimates.

Comparison	***t* = t x μ**	**μ**	**Mean**	**HPD90Lo**	**HPD90Hi**
Coastal and West Beringia	8.02	2.20E-06	3,637,665	1,696,368	5,905,536
Coastal and Interior Alaska	3.55	2.20E-06	1,610,188	612,325	2,608,052
Coastal and North Slope	12.57	2.20E-06	5,703,696	1,530,813	11,600,160
Coastal and West Beringia	8.02	1.47E-06	5,454,299	2,543,526	8,854,735
Coastal and Interior Alaska	3.55	1.47E-06	2,414,310	918,118	3,910,501
Coastal and North Slope	12.58	1.47E-06	5,454,299	2,295,294	17,393,230

Table 5 - Estimates of θ from the three Isolation with Migration comparisons.

Values for the parameter θ are given for the three pairwise Isolation with Migration comparisons. Only θ for the contemporary populations is shown.

Comparison	Coastal θ	HPD90Lo	HPD90Hi	Other θ	HPD90Lo	HPD90Hi
Coastal and West Beringia	5.68	3.04	10.14	0.53	0.15	1.23
Coastal and Interior	4.10	2.02	7.61	0.62	0.28	1.30
Coastal and North Slope	4.88	2.88	7.94	0.22	0.05	0.69

Chapter 4: Conclusion

Several general conclusions can be drawn from the information produced as part of this thesis research and reported in the two chapters on blackfish genetics. Blackfish, as expected from their low dispersal ability and tolerance to hypoxic and cold conditions, have a high degree of genetic variability. The observed genetic variability permitted the exploration of several hypotheses concerning the history and connections between living populations of blackfish.

A central question that motivated this thesis study is: "Are blackfish composed of taxonomically distinct populations?" Originally this study was spurred on by the observed karyotype differences between North Slope and Yukon River blackfish. I found that while genetic differentiation was supported for up to five geographic regions, the previously described *D. admirabilis* did not fit widely used criteria for species delineation from a mitochondrial perspective. On the other hand, the analyses presented here are silent regarding any current reproductive isolation of the Amguema River population from other blackfish populations. All fish sampled from the North Slope carried haplotypes that represent a monophyletic mitochondrial lineage. Blackfish from the North Slope were nearly all homozygous for a private RAG1 allele, but they had common S7 alleles. These lines of evidence strongly support the distinctiveness of North Slope blackfish as a taxonomic entity. The very low amount of estimated migration will permit a large amount of genetic differentiation, and due to the geographic isolation and disjunct distribution of North Slope blackfish there is likely no ongoing migration occurring. Geography and the genetic evidence reported here suggest that North Slope

blackfish are an example of incipient speciation. Elsewhere genetic structuring is clear in separating blackfish into geographic areas along likely biogeographic barriers.

Another core question that motivated this study was: "Is there evidence for multiple aquatic refugia within the greater Beringian refugium?" The high degree of genetic variability of blackfish has resulted from isolation in multiple Beringian refugia. It is clear that Beringia contains multiple aquatic refugia where blackfish persisted during times of paleoclimatic instability during the Pleistocene. West Beringia and Coastal Alaska (Yukon and Kuskokwim drainages adjacent to the former Bering land bridge) are similar in genetic composition, and probably part of a greater Bering land bridge refugium that utilized the wetter exposed continental shelf during times of glacial maxima. Parts of the Tanana River drainage and the Kuskokwim River drainage upstream of the Kuskokwim Mountains formed another aquatic refugium during the Pleistocene. Although the North Slope of Alaska is thought to have been an arid, cold landscape during glacial advances, the genetic evidence suggests that it was able to support Blackfish for a long time. Finally, Unalakleet and Nome appear to also have been a separate glacial refugium.

Finally, a third line of inquiry guiding this study was the Bering land bridge with such questions as: "How did the Bering land bridge influence intercontinental aquatic interchange? Was the Bering land bridge a conduit to migration? Were Blackfish able to traverse it at all?" The high degree of genetic similarity across the Bering land bridge informed hypotheses regarding the Bering land bridge's influence on intercontinental aquatic interchange. Of the five groups of sampled populations defined, West Beringia

and Coastal Alaska were the most similar in genetic composition. Furthermore, estimates of migration rates between the two areas indicate a substantial level of exchange. Based on these results, it appears that the Bering land bridge was conducive to intercontinental interchange for blackfish during times of glacial advances. The Bering Sea opened and closed this gateway numerous times in a cyclical fashion, as migration of blackfish across the Bering Sea is extremely unlikely. Despite intermittent migration due to presence of the Bering land bridge followed by the Bering Sea, relatively high levels of migration are estimated across the Bering Strait.

This research highlights the value of blackfish studies as a model system for aquatic Beringian biogeography. Further research including the type locality of *D. delicatissima,* Kotzebue Sound, and the Bristol Bay region with additional markers would complement the information produced by this work. However, based on the evidence collected to date, it seems unlikely that additional research will substantially change the general conclusions that were reached and reported in this study.